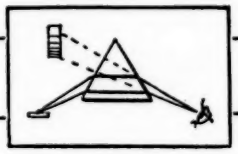
Journal of the OPTICAL SOCIETY *of* AMERICA

VOLUME 28
No. 3



MARCH
1938

Grating Monochromator for the Schumann Region . . . RICHARD F. BAKER 55

Calibration Wave-Lengths for Infra-Red Spectrometers . . . . . P. E. SHEARIN AND E. K. PLYLER 61

Investigations on the Intrinsic Properties of the Color Domain . . . . . LUDWIK SILBERSTEIN 63

Study of Harting's Criterion for Complete Achromatism . . FRED H. PERRIN 86

Published monthly
at Prince and Lemon Sts., Lancaster, Pa.
by the
AMERICAN INSTITUTE OF PHYSICS
Incorporated
175 Fifth Avenue, New York, N. Y.
FOR AND UNDER THE EDITORIAL SUPERVISION OF
THE OPTICAL SOCIETY OF AMERICA

THE OPTICAL SOCIETY OF AMERICA

Officers of the Society

R. C. GIBBS, *President*
Cornell University
Ithaca, New York

K. S. GIBSON, *Vice President*
National Bureau of Standards
Washington, D. C.

L. B. TUCKERMAN, *Secretary*
National Bureau of Standards
Washington, D. C.

HENRY F. KURTZ, *Treasurer*
Bausch & Lomb Optical Company
Rochester, New York

The Board of Directors consists of the above officers; the Editor of the Journal; the Past President; and four elected members.

A. C. HARDY, *Past President*
Massachusetts Institute of Technology
Cambridge, Massachusetts

F. K. RICHTMYER, *Editor*
Cornell University
Ithaca, New York

Elected Members of the Board of Directors

BRIAN O'BRIEN
University of Rochester
Rochester, New York

H. P. GAGE
Corning Glass Works
Corning, New York

WILLIAM F. MEGGERS
National Bureau of Standards
Washington, D. C.

G. R. HARRISON
Massachusetts Institute of Technology
Cambridge, Mass.

Rochester, New York, Section of the Optical Society of America

B. NOYES, *President* G. FASSIN, *Vice President* F. M. E. HOLMES, *Secretary* J. S. CAMPBELL, *Treasurer*

Councillors: H. F. KURTZ, F. C. FAIRBANKS, A. L. SCHOEN, J. H. WEBB, L. A. DUBRIDGE (*ex officio*)

Corporation Members of the Optical Society of America

American Optical Company, Southbridge, Massachusetts
Bausch & Lomb Optical Company, Rochester, New York
Bell Telephone Laboratories, Inc., New York, New York
Corning Glass Works, Corning, New York
Gaertner Scientific Corporation, Chicago, Illinois
Interchemical Corporation, New York, New York
Keuffel & Esser Company, Hoboken, New Jersey
Spencer Lens Company, Buffalo, New York
Arthur H. Thomas Company, Philadelphia, Pennsylvania

THE OPTICAL SOCIETY OF AMERICA was organized in 1916 "*to increase and diffuse the knowledge of optics in all its branches, pure and applied, to promote the mutual interests of investigators of optical problems, of designers, manufacturers and users of optical instruments and apparatus of all kinds and to encourage cooperation among them.*" It aims to serve the interests of all who are engaged in any branch of optics from fundamental research to the manufacture of optical goods.

MEMBERSHIP

Regular Members: "Any person who has contributed materially to the advancement of optics is eligible for regular membership."

Associate Members: "Any person who is interested in optics is eligible for associate membership.

Corporate Members: "Any corporation desirous of encouraging the work of the Society is eligible for corporation membership."

The Society cordially invites to membership all who are interested in any branch of optics, either in research, in instruction, in optical or illuminating engineering, in the manufacture and distribution of optical goods of all kinds or in physiological and medical optics.

ANNUAL DUES

Regular members, $7.50; Associate members, $5.00; Corporate members, $50.00. These dues include subscription to the Society's official journal, the *Journal of the Optical Society of America*.

MEETINGS

The Society holds two meetings per year: An annual meeting in the autumn and a midwinter meeting jointly with the American Physical Society.

For further information concerning the Optical Society of America address the Secretary, Dr. L. B. Tuckerman, National Bureau of Standards, Washington, D. C.

Journal of the
OPTICAL SOCIETY
of AMERICA

VOLUME 28 MARCH, 1938 NUMBER 3

A Grating Monochromator for the Schumann Region

RICHARD F. BAKER*
Department of Physics, University of Rochester, Rochester, N. Y.
(Received January 17, 1938)

A vacuum grating monochromator is described, suitable for photoelectric research in the Schumann region. A hydrogen discharge tube serves as a light source, and radiation intensities are measured with a potassium photo-cell and fluorescent filter. Photo-currents are measured with a Compton electrometer, using the steady deflection method. The photoelectric yield of cadmium has been measured from the threshold to 1000A, and curves are given showing a maximum for the surface effect and the onset of the volume effect.

INTRODUCTION

MOST researches on the photoelectric effect have been confined to the visible and that portion of the ultraviolet spectrum which is accessible with a quartz prism. Sabine[1] confirmed the validity of the Einstein law in the Schumann region, and Kenty,[2] using the filtered radiation from a rare gas discharge reported evidence for very high yields below 1000A. Aside from this work, no quantitative measurements on the photoelectric effect exist for wave-lengths less than 1900A.

As Lyman[3] and, more recently, Bomke[4] have pointed out, the Schumann region is one of great interest and importance for photochemistry, photoconductivity and photoelectricity, because the higher energy quanta initiate effects which are not observed in the visible and quartz ultraviolet. The predictions of theory for the photoelectric effect are especially interesting. The observed photoelectric currents in the vicinity of the threshold are the result of surface absorption, and this surface yield should pass through a maximum. Quantum theory[5] demands a second rise on the high frequency side of the maximum, due to excitation of electrons from the top occupied energy zone within the body of the metal. The threshold for this volume effect is usually in the vacuum ultraviolet. Volume thresholds have never been observed for any metal other than potassium, because of their inaccessibility.

The purpose of this paper is to describe a grating monochromator for photoelectric work in the Schumann region, and to present some preliminary results obtained with it.

GENERAL DESCRIPTION

A grating instrument, used at near grazing incidence, was chosen as a monochromator for the Schumann region. This choice was dictated by

* Now at University of Minnesota.
[1] P. E. Sabine, Phys. Rev. **9**, 210 (1917).
[2] Kenty, Phys. Rev. **44**, 891 (1934).
[3] Lyman, *The Spectroscopy of the Extreme Ultraviolet*, (1928), p. 124.
[4] Bomke, Physik. Zeits. **37**, 222 (1936).
[5] Tamm and Schubin, Zeits. f. Physik **68**, 97 (1931).

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

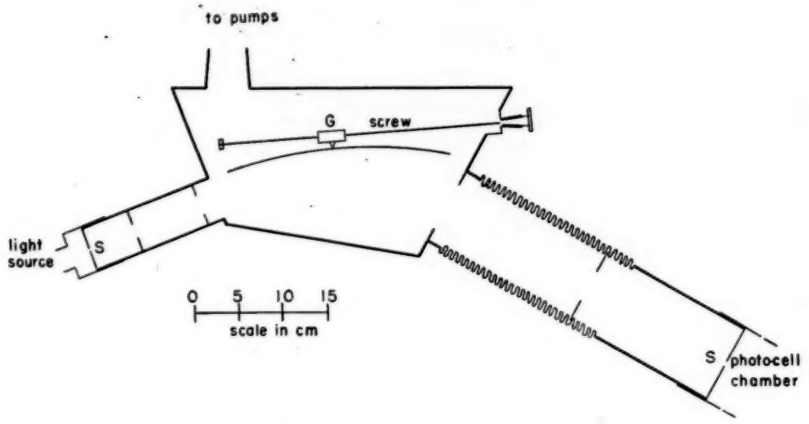


FIG. 1. Plan of vacuum spectrograph, drawn to scale. *S* denotes slits, *G*, grating. Sylphon bellows control is not shown.

the need for high dispersion and high reflecting power. The concave grating was of one meter focal length, 569 lines per mm, ruled on glass and coated with an evaporated aluminum film to increase the reflectivity. It was made available through the courtesy of the Eastman Kodak Research Laboratory.

The grating was housed in a brass chamber, 15 cm high, the plan of which is shown in Fig. 1. The walls were screwed and soldered together. Brass tubes 5 cm and 11 cm in diameter were soldered into opposite sides of the box, as shown in the figure, and served as incidence and exit tubes.

The movable element in this monochromator is the grating, since it was not practical to move either the entrance or exit slit. The grating was mounted on a carriage which slid along two V-shaped grooves in the base plate of the chamber. These grooves were accurately machined to the radius of the Rowland circle. Motion could be communicated to the grating carriage by means of a floating nut moving on a long fixed screw which could be operated from the outside by means of a greased cone joint. A millimeter scale was engraved on the base plate, with a pointer on the grating carriage to serve as a wave-length scale. The grating mount was provided with the usual adjustment screws for adjusting the grating to tangency with the Rowland circle. A removable section in the chamber top made the grating accessible, and a plate glass window was provided through which the scale position could be read.

The monochromator as a whole stood on three legs which were screwed to a heavy brass plate which in turn was clamped to a table.

A sylphon bellows in the exit tube served the purpose of extending the available wave-length range beyond that furnished by grating movement. This could easily be done by moving the bellows so as to locate the exit slit at a different point on the Rowland circle. A segment of a circle of proper radius of curvature was engraved on the brass base plate, and the exit tube provided with a pointer to facilitate this adjustment. The fundamental range of the instrument is from 1000A to 2500A, and with the aid of the bellows a range of 500A to 3000A is available. The grating movement for the fundamental range is about 10 cm, which with a screw of 25 threads per inch, amounts to about 95 complete turns. The angle of incidence varied from 74°13′ at one extreme position of the grating to 68°45′ at the other extreme.

The entrance slit was very simple in design, consisting of one fixed jaw, and one movable jaw, sliding in a V-shaped groove, manually adjusted and held in place with a brass spring. It was mounted on one end of a brass tube which fitted snugly into the entrance tube. The slit was 5 mm long and usually about 0.05 mm wide. A number of fixed-width exit slits were made, each 10 cm long, which fitted over the end of the exit tube. The exit slit used in the experiments reported

here admitted a band of radiation approximately 75A wide, at a mean position of the grating.

The monochromator was evacuated with two brass oil diffusion pumps of the umbrella type, backed by a Megavac, the pump connection being made with a glass to brass waxed cone joint.

Pressures were measured with a Pirani resistance gauge, giving a continuous reading from 1 mm of Hg to 10^{-4} mm of Hg.

Astigmatism

In determining photoelectric sensitivity curves, the radiation intensity is usually measured with a thermopile or radiometer. Either the monochromator or the photo-cell must then be movable, since it is essential to measure the intensity and the photo-current as nearly simultaneously as possible. In this monochromator advantage is taken of the astigmatism to avoid introducing any movable elements. The general expression for length of the line image for a concave grating is:

$$L=2\left[z\left(\sin^2\theta_2+\frac{\cos\theta_2}{\cos\theta_1}\cdot\sin^2\theta_1\right)+s\cdot\frac{\cos\theta_2}{\cos\theta_1}\right],^{6} \quad (1)$$

where z = half-length of ruling,
s = half-length of entrance slit,
θ_1, θ_2 = incidence, excidence angles.

Values of z, s, θ_1, θ_2 appropriate to this monochromator lead to a line length of 11 cm.

The photo-cell and the thermopile substitute to be described below are mounted vertically one over the other, behind the exit slit, and approximately half of the spectral image falls on each. Thus simultaneous measurements of photo-current and intensity are possible. The apertures were approximately equal, though no especial attention was paid to this point, as relative intensities only were of interest. For the same reason a slight inhomogeneity of the line over its length would be unimportant.

Dispersion

Using a concave grating mounted as described above, it can be shown that to a first approximation the dispersion is constant. If Δs denote the distance through which the grating moves along the arc of a Rowland circle and $\Delta\lambda$ denote the resultant shift in wave-length of the radiation incident on the exit slit, then

$$\Delta\lambda/\Delta s=(d/2\rho)[\cos i+\cos r]. \quad (2)$$

The dispersion is constant to the approximation that the bracketed term on the right-hand side of Eq. (2) is constant over the range of i and r. Over the range of values for this instrument the dispersion varies by 1.2 percent of its mean value, which is 14.9A per mm.

Light Source

A light source for photoelectric work must possess adequate intensity, stability, and some measure of reproducibility. In addition it is very convenient to have a line emission spectrum, rather than a continuum. These requirements, so well met by the quartz mercury arc in the ultraviolet and visible, are difficult to satisfy in the Schumann region. Most sources of adequate intensity such as the vacuum spark are far from stable. After much experimenting, a first approximation to an ideal source was found in the high voltage uncondensed hydrogen discharge through a quartz capillary, operated at a power input of the order of 10 kilowatts. Both electrodes and capillary were water cooled. One electrode was waxed to the entrance tube, the discharge being end-on to the slit. Hydrogen was admitted to the tube through a leak valve and passed through the slit and monochromator, being pumped out by the main pumps. Power was supplied from a 15 kw 10,000-volt transformer operating on the 110-volt a.c. line. The circuit included a ballast resistance in series with the primary and an ammeter in the secondary circuit. Tank hydrogen was found satisfactory. The current through the tube was kept at about 2 amp. As long as pressure and current were maintained constant this source proved very constant and reproducible. The spectrum is, however, not a line spectrum, but above 1600A, a continuum and below 1600A, a many line or band spectrum, characteristic of the hydrogen molecule.

Experimental Photo-Cell

The experimental photo-cell approximated the black-body type, as shown in Fig. 2. Due to the

[6] Meyer, *The Diffraction of Light, X-Rays, and Material Particles*, Appendix E, p. 438.

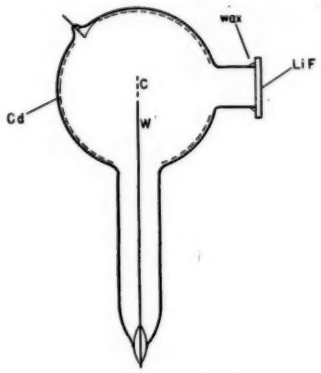


FIG. 2. Experimental photo-cell. *C* denotes collector, a loop of tungsten wire, spot-welded through nickel to the main tungsten lead *W*.

low radiation intensities available in the Schumann region, a cell of this type possesses marked advantage over the filament cathode type of cell, as more of the incident radiation is useful photoelectrically. Cadmium was chosen as the element to be investigated, because of its conveniently located threshold at 3150A and the ease with which it is distilled. The cadmium was distilled several times in a high vacuum before being introduced into the cell. The window, a synthetic LiF crystal whose transmission limit is ~1000A, was waxed on with beeswax and rosin, and kept clean during distillation by a thin piece of mica. The cell was sealed off the pumps at a pressure less than 1×10^{-6} mm of Hg, and mounted in the monochromator. No attempt was made to work with extremely clean surfaces in these preliminary tests.

INTENSITY MEASUREMENTS

Radiation intensity measurements, troublesome at best, are notoriously difficult in the Schumann region because of the low intensities. The usual instruments, thermopile-galvanometer, bolometer, radiometer, etc., are far too insensitive.

A technique developed by Dubouloz[7] and others[8, 9] seemed most promising and was investigated. The scheme is to convert a photo-cell into a nonselective detector of radiation by coating the window with a fluorescent dye. For any incident wave-length (less than a certain value) the photoelectric surface is illuminated always by the same (fluorescent) wave-length band. The working assumption is that the quantum efficiency of fluorescence is independent of frequency. The resultant photo-current then is proportional to the number of incident quanta and we have,

$$I_1/I_2=(i_1\nu_1)/(i_2\nu_2), \tag{3}$$

where I_1, I_2 are radiation intensities in energy units, of frequencies ν_1 and ν_2, which give rise through fluorescence to photo-currents i_1 and i_2.

A potassium in Pyrex photo-cell was used, the window of which was coated with the Eastman chemical di-ethyl-di-hydrocollidine-dicarboxylate. This dye is one used to sensitize Eastman ultraviolet plates to light of short wavelength. Its fluorescent band is centered at 4200A. The potassium cell was designed for maximum sensitivity to light of 4000–4500A wave-length, being sensitized by glow discharge in hydrogen and filled with helium at a low pressure. The dye was applied from benzol solution with a brush and allowed to dry, leaving a thin crystalline layer over the window.

It is of course essential that the photo-cell itself exhibit a linear response to radiation of varying intensity, and this was found to be the case for low intensities.

Measurements taken with a quartz monochromator, mercury arc and thermopile in the near violet, together with a plausible assumption as to the location of the Fowler threshold for K, made possible a calculation of the quantum efficiency of the dye. It was found to be of the order of 1/20, i.e., one quantum of fluorescent light was emitted on the average for each 20 quanta of incident light of short wave-length.

The vacuum thermopile used was of the type designed by Mendenhall, used with a Zernicke Zc galvanometer. For an incident energy sufficient to give a 1 mm deflection on the thermopile the K photo-cell current was 1.5×10^{-12} amp., or an electrometer deflection of 1500 mm at full sensitivity. The minimum energy detectable is thus only $\sim10^{-3}$ of that required to give a thermopile deflection of 1 mm.

[7] P. Dubouloz, Rev. d'optique **12**, 114 (1933).
[8] A. Chevallier and P. Dubouloz, Comptes rendus **194**, 174 (1932).
[9] G. Déjardin and (Miss) Schwegler, Rev. d'optique **13**, 313 (1934).

Current Measurement

The two photo-cells were, as mentioned above, mounted one above the other directly behind the exit slit in a brass cylinder with removable top which was waxed to the exit tube, being evacuated through the exit slit. The photo-cell leads were sealed through holes drilled in the top and back of the cylinder. From these seals the collector leads went through a manually controlled switch for shifting cell connections to the quadrant of a Compton electrometer. The usual shielding precautions were observed. The steady deflection method was used, with S. S. White resistors, and sensitivities ranged from 4000 to 15,000 mm/volt, at a scale distance of two meters.

Scattered Light

The problem of scattered light demanded special consideration. Radiation intensity measurements in particular were susceptible to error from this cause, inasmuch as the potassium cell was highly sensitive to the visible as well as the ultraviolet. The entrance and exit tubes were diaphragmed and the interior of the grating chamber painted an optical black. Furthermore two diaphragms were mounted on the grating carriage to reduce reflection from the mounting. Enough light was scattered from the ruled area, however, to cause serious error.

Obviously a filter opaque to the visible and quartz ultraviolet but transparent to the Schumann region was needed. Since thin sodium layers have this property a sodium filter was constructed, the sodium being distilled into a thin Pyrex cell closed at either end with LiF windows. Visual tests were made with a hand spectrometer and mercury arc during the distillation to help fix the optimum film thickness. This cell was sealed off and proved very satisfactory in reducing the amount of scattered light, though the cut-off was not as sharp as anticipated and the filter did transmit a little radiation between 2000A and 3000A. This radiation could not be eliminated, but was corrected for by the expedient of taking two sets of data, one set with fused quartz filters in front of the cells, and one set without filters. The photo-current due to the radiation in the Schumann region for which the grating was set is then given by the difference in deflections at that setting.

Table I.

λA	Electrons/Quantum
3130	0.008×10^{-4}
2650	0.50 " "
2480	0.64 " "
1762	1.3 " "
1608	2.5 " "
1452	2.5 " "
1300	2.8 " "
1149	46.3 " "
1000	200.0 " "

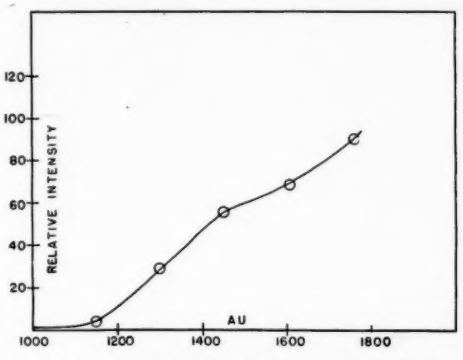


Fig. 3. Intensity distribution of a hydrogen discharge, plotted in arbitrary energy units, *versus* wave-length. Scattered light intensities have been subtracted.

Results

Figure 3 shows the intensity distribution of the hydrogen discharge after modification by passage through a LiF crystal. This curve has been corrected for scattered radiation.

In Table I are tabulated data on the photoelectric yield of cadmium, expressed in electrons per quantum. The logarithm of the yield has been plotted in Fig. 4, against the wave-length. The yields near the threshold are of the same order of magnitude as those found for other metals by other workers. Kenty,[2] using the resonance line of argon at 1066A has found a yield of 60×10^{-4} electrons/quantum for a nickel surface, and 70×10^{-4} electrons/quantum for tungsten, which may be compared with the values in Table I.

The photo-current per unit incident energy has been plotted in Fig. 5, showing the slight surface maximum almost masked by the rise of the volume effect. The ratio of the frequency at the

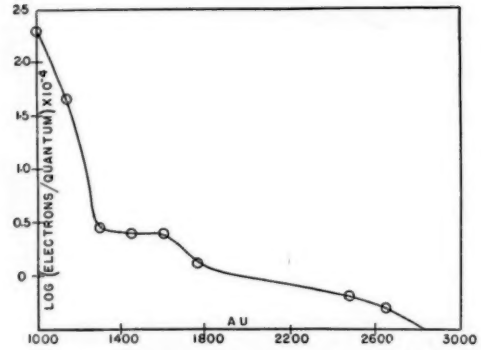


FIG. 4. Photoelectric yield of cadmium *versus* wave-length. Ordinates are the log of the yields given in Table I.

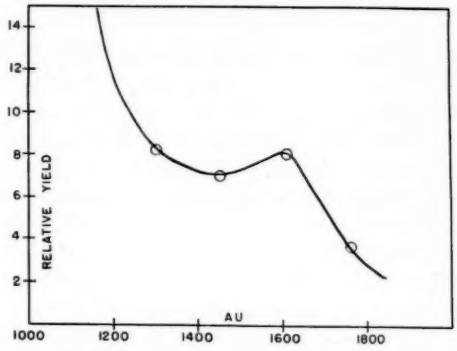


FIG. 5. Photoelectric yield, expressed in terms of photocurrent per unit incident energy, *versus* wave-length, showing surface effect maximum, and rise of volume effect. Yields are in arbitrary units with above dimensions.

maximum to the threshold frequency is nearly 2. Extrapolation of the short wave-length side of Fig. 5 to the axis gives a value of about 1500A (8.2 volts) for the volume threshold. Kenty[2] has suggested that the surface maxima for the surfaces he studied must lie on the short wave side of 584A, his shortest available wave-length. In view of the results described here for cadmium, it seems possible that for his surfaces the maxima may lie between 1066A and 2000A, where no data were taken, the rise below 1066A being caused by the onset of the volume effect. The latter may have masked the surface peak.

In conclusion the author wishes to express his indebtedness to Professor L. A. DuBridge, under whose direction the work was undertaken, and to other members of the staff for helpful suggestions. The work was financed in part by a grant-in-aid from the National Research Council.

Calibration Wave-Lengths for Infra-Red Spectrometers

P. E. SHEARIN AND E. K. PLYLER
Physics Department, University of North Carolina, Chapel Hill, North Carolina
(Received January 4, 1938)

About 20 wave-lengths which have been measured on grating spectrometers have been given for the calibration of prism spectrometers in region from about 2μ to 24μ. A number of lines due to H_2O and CO_2 can be normally observed due to atmospheric absorption. Since not many sharp bands have been measured on grating spectrometers from 7μ to 13μ, several sharp bands in ethyl alcohol and amyl alcohol in the liquid state have been measured in this region. The values found in the present work check well with those previously found by Weniger except those in the 9μ and 10μ region. Here some of the bands differ by about 0.2μ from those found in this work. This difference is probably due to the fact that Weniger did not have the more accurate values of the index of refraction of rocksalt due to Paschen at his disposal.

THE index of refraction of rocksalt has been measured by Paschen[1] over the region from 0.5893μ to 15.9116μ. The index of refraction of other prism materials is also known in the regions where they are not absorbing in the infra-red.

If a prism mounting of the Wadsworth type is used it is not difficult readily to translate the angular reading to wave-lengths.

Cross[2] has recently calculated the index of refraction of rocksalt for the temperature of 25°C using Paschen's original values. These values are primarily suited for 60° prisms. He finds that a change of temperature of 4°C from 25°C may result in an error of 0.01μ in wave-length. The angle of minimum deviation is a function of temperature and when the prism is correctly mounted at one temperature, it will not be mounted at minimum deviation for other temperatures. The calibration will also be slightly affected by any change in position of the source of radiation. With these and other errors which may be present it seems desirable to have a separate and independent method of check for the wave-length calibration of a prism instrument.

By the use of rotational lines and sharp bands as measured by a grating spectrometer we have a means of checking calibration for any kind of prism used in the infra-red region. As the grating spectrometer has a resolution much greater than the prism spectrometer, the errors in the measured wave-lengths are much less than the effective slit width of the prism instrument. The purpose of this paper is to give a table of wave-lengths suitable for calibration and a brief discussion of their use and limitations.

The infra-red absorption bands of a number of gases have been measured by the use of a grating spectrometer. When the band has a zero branch, it will serve as a good calibration point. The CO_2 band at 14.97μ and the N_2O band at 17.0μ are examples of this type. In the double type of band, having only P and R branches, the center often can be found on account of its being a region of less absorption if the frequency interval between the first lines on each side is sufficiently great. The HCl band at 3.46μ and H_2O at 6.26μ are examples of this type of band. The coefficient of absorption varies for the different gases and the original papers should be consulted for details as to cell thickness and pressure. If the spectrometer has sufficient resolving power to separate the individual lines of a band, the lines are well suited for calibration.

A cell is usually necessary for the study of the absorption bands of gases, but on account of the presence of H_2O and CO_2 in the optical path of the spectrometer their spectra can usually be observed without the use of a cell. The bands of these substances are well suited for calibration as they fall in various regions of the infra-red. A list of bands for different substances is given in Table I for the region of 2μ to 24μ.

Since usually it is easier to make liquid cells than to generate gas for a gas cell it was thought

[1] Paschen, Ann. d. Physik **26**, 120 (1908).
[2] Cross, Rev. Sci. Inst. **4**, 197 (1933).

advisable to measure the wave-lengths of some sharp absorption bands of common liquids obtainable in any laboratory. On looking over the infra-red data for various liquids it was found that very few measurements have been made by means of grating spectrometers, for liquids with sharp bands. After studying the absorption bands of Weniger[3] for a number of organic liquids it appeared that ethyl alcohol and amyl alcohol would serve our purpose well. These are found to have sharp absorption bands extending from 2μ to 13μ. As there are several good calibration points in the shorter wave-length region the present study extends from about 7μ to 13μ. Measurements in this region were made in a grating spectrometer having 2060 lines/inch and with an effective slit width at 9μ of about 2 waves/cm. The cells containing the liquid were made by pressing a drop of the liquid between two polished rocksalt plates (fluorite windows were used in the 7μ region). The resulting cell thickness was of the order of 0.01 mm. It was found that the position of the bands did not change with a change in cell thickness as might possibly happen with a band that is not symmetrical.

TABLE I. *Wave-lengths of substances as measured on grating spectrometer suitable for calibration of prism spectrometers.*

SUBSTANCE	WAVE-LENGTH	REMARKS
HCl[1]	1.764μ	center of band
	3.465	center of band
H_2O[2]	2.673	zero branch
	6.263	center of band
CO_2[3]	4.255	center of band
	14.97	zero branch
C_2H_5OH[4]	6.945	maximum absorption region
	7.169	maximum absorption region
	7.466	maximum absorption region
	9.074	maximum absorption region
	9.421	maximum absorption region
	11.220	maximum absorption region
$(C_5H_{11})OH$[4]	9.369	maximum absorption region
chain type	9.832	maximum absorption region
	11.801	maximum absorption region
	12.012	maximum absorption region
	12.961	maximum absorption region
H_2O[5]	16.190	rotational lines
	16.893	rotational lines
	19.023	rotational lines
	21.12	rotational lines
	23.81	rotational lines

[1] Meyer and Levin, Phys. Rev. **34**, 44 (1929).
[2] Sleator and Plyler, Phys. Rev. **37**, 1493 (1931).
[3] Martin and Barker, Phys. Rev. **41**, 291 (1932).
[4] Measured in present work.
[5] Weber and Randall, Phys. Rev. **40**, 835 (1932).

[3] Weniger, Phys. Rev. **31**, 388 (1910).

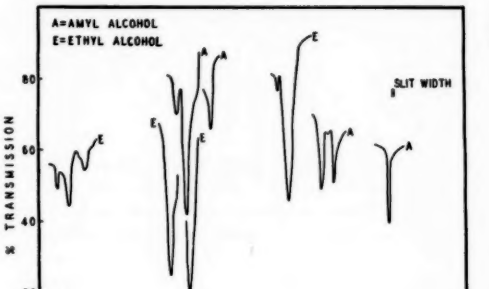


FIG. 1. The infra-red absorption of ethyl and amyl alcohols from 7μ to 13μ. Cell thickness about 0.01 mm.

It will be noted that the positions of bands in the region of 7μ and in the region of 11μ and 12μ check almost exactly with those of Weniger in spite of the fact that our resolving power was 5 to 10 times that of his. This shows that the position of sharp symmetrical bands like those found in the alcohols is independent of the resolving power of the spectrometer. In the region of 9μ obviously Weniger's calibration was not accurate. Paschen[1] has pointed out that Langley's values for the index of refraction of rocksalt (used by Weniger for calibration) are in error in the region of 7μ. Since his values did not extend beyond 7μ it was necessary for Weniger to extrapolate his curve to the region of 9μ, and, therefore, one might expect errors in this region.

In Fig. 1 are shown the absorption bands as found in the region 7μ to 13μ for ethyl and amyl alcohol. The sharp bands at 6.945, 7.169, 7.466, 9.074, 9.421, and 11.220μ for ethyl alcohol are well suited for calibration purposes. Also the bands at 9.369, 9.832, 11.801, 12.012, and 12.961μ in amyl alcohol are sharp and furnish additional calibration points which are needed, as not many studies have been made by the aid of grating spectrometers in this part of the spectrum.

Investigations on the Intrinsic Properties of the Color Domain*

LUDWIK SILBERSTEIN
Eastman Kodak Research Laboratories, Rochester, New York
(Received January 12, 1938)

I. INTRODUCTION

THE manifold of all colors perceivable by a normal eye is, when properly ordered,[1] of three dimensions, or briefly a *threefold*. For, in order to fix a color, three independent data, such as, for example, the attributes of color revealed by introspection, hue, saturation, and brilliance, are necessary and sufficient. For the purpose of quantitative discussion the psychophysical correlate of brilliance must be used. This correlate is brightness and is measured by the methods of heterochromatic photometry.

Before discussing the complete, three-dimensional domain of colors it seems advisable to consider a sub-domain of this threefold, *viz.*, the totality of colors of *equal brightness*, which, without loss to generality, may be taken as *unit* brightness. With this deliberate limitation the color domain to be first contemplated will be a manifold of two dimensions only, or a *twofold*.

II. THE TWOFOLD OF EQUIBRIGHT COLORS

Since the times of Newton this twofold has, with everybody's tacit consent, been habitually represented by a (portion of a) *Euclidean plane*, a one-to-one correspondence being established between the particular colors and the points of such a plane and also between sets of colors obtainable by additive mixing of two colors and the straight lines, i.e., the geodesics of the plane. In adopting this procedure it was tacitly assumed that the twofold of colors has, intrinsically, the same structural properties as that geometrical twofold, the Euclidean plane. We have here in mind not merely the *metrical*, but even the broader, *viz.* the *projective* properties of a plane, treated as a sub-manifold of the projective three-dimensional space.

Now, it is by no means a settled question whether such an identification or correlation of the color twofold and the geometrical plane is legitimate or not.[2] In the latter case the familiar representations, as e.g., the usual color triangle, are manifestly objectionable, since they carry with them unwarranted distortions of the color twofold.

Our first object is to show how this fundamental question can be decided experimentally, either in the affirmative or the negative.

To begin with, this question of representability or nonrepresentability of the color twofold by a portion of the Euclidean, more generally, of the projective plane can be readily seen to be reducible to the more definite question:

Is this two-dimensional color domain a *Desarguesian* twofold or not? That is to say, does the theorem of Desargues hold in the color twofold when the geometrical "points" and "lines" in its usual enunciation (to be given presently) are replaced by single colors, all of equal brightness, and by shortest color lines?[3]

To repeat it, this fundamental question is not a matter of opinion (or "theory"), but must and can be decided only empirically. We may say beforehand that the answer to this question is most likely to turn out in the negative.

In approaching the discussion of this question it will be helpful to introduce here a conceptual entity which is roughly, but to all probability not strictly, analogous to the geometrical

* Communication No. 658 from the Kodak Research Laboratories.

[1] A manifold of elements, whether discrete or continuous, taken in itself, is devoid not only of metrical properties, but also of dimensionality. The latter is impressed upon the manifold, more or less artificially or arbitrarily, by *ordering* its elements into sets or sub-manifolds in some way or other.

[2] Doubts pertaining to this question were incidentally expressed by previous writers on this subject. Thus, e.g., the much regretted L. T. Troland says in his *Psychophysiology* (Vol. I, p. 250) with much emphasis, although without attempting a mathematical treatment: "This space of the color system does not necessarily have the same mathematical properties as does geometrical or physical space. In other words, it may not be Euclidean," and he draws even the reader's attention to the fact that there is no particular reason why the ratio of "the saturation of a cycle of equally saturated chromas and the number of just noticeable hue steps in such a cycle should be equal to 2π" (the ratio of circumference to radius in a Euclidean circle).

[3] I.e., one-dimensional sets of equibright colors, the passage along which from one color A to another B comprises the *least* number of just perceptible "steps" or differences. (Cf. *infra.*)

"straight line," and which will be called a *suite of colors*. This will be defined presently in terms of a pair of colors and a certain process (*viz.* "mixing" or "blending") to be performed upon them.

In what follows Roman capitals, A, B, and so on, will be used for particular colors and small Greek letters, α, β, etc. or small Romans, x, y, etc. (denoting ordinary, real, positive numbers) for their brightness. These will be used as "factors" of A, B, etc.

Let the common brightness of the colors of the contemplated twofold be taken as *unit*. Consistently, let A, B, C, $\cdots$ be each a color at unit brightness. A color A at brightness α (to be mixed with another color) will be written αA.

Now, consider two such colors, A and B, supposed distinct from each other, and let these be compounded by an operation generally known under the familiar name of additive "mixing" or "blending."

The two colors, the former of brightness α, and the latter of brightness β, may be supposed here to be blended optically. We can safely assume that the brightness of the resultant color P is *the sum* of the brightnesses of the components. Thus, if we make $\alpha+\beta=1$, P will be a color of the contemplated twofold. It may be called "the sum" of αA and βB and written

$$P=\alpha A+\beta B, \qquad \alpha+\beta=1, \tag{1}$$

or, to avoid the explicit mention of the relation between α and β,

$$P=(uA+vB)/(u+v) \tag{1a}$$

where the (positive) numbers u, v are unrestricted, independent.

Since the two component colors play, in the process of blending, the same role, this "sum" is obviously *commutative*,

$$\alpha A + \beta B = \beta B + \alpha A.$$

It is also, in the case of blending together three or more colors, *associative*, i.e.,

$$(\alpha A+\beta B)+\gamma C=\alpha A+(\beta B+\gamma C).$$

To see this, the reader has only to express the several terms of this equation in colors and their blending. In fine, then, the color "sum" has all the properties of the ordinary algebraic sum.

If, in (1), α and β $(=1-\alpha)$ are varied continuously, we generate a one-dimensional manifold of colors P, all within the twofold, ranging from pure A $(\alpha=1)$ to pure B $(\alpha=0, \beta=1)$.

Such a manifold will be called a *suite* of colors, extending from A to B, and may be briefly denoted by AB.

The *extension* or protraction of such a suite (more properly of the segment of a suite) can be defined as follows:

If C be a color, always of unit brightness, such that

$$B=\gamma C+(1-\gamma)A,$$

where γ is again a positive number, less than unity, then the color C will be said to belong to or to "lie on" the suite AB, *viz.* "beyond" B. In the same manner the segment AB can be extended beyond A.

The last equation can also be written, in a purely formal way,

$$C=(1/\gamma)[B-(1-\gamma)A],$$

but one must always remember that "subtraction" of colors (usually associated with absorption by a dye) is entirely foreign to our circle of ideas, nay meaningless, and that the presence of a negative coefficient, like $-(1-\gamma)/\gamma$, in a binomial of colors signifies only that the algebraic sum or the resultant color is *beyond* one of the addends.

Such an extension of a color suite AB cannot, of course, be continued indefinitely, but only up to certain limits, say S and T, the former beyond A and the latter beyond B, which may be called *extreme* or *boundary* colors. In other words, whatever the pair A, B, it may be considered as an experimental fact that there is a pair of colors S, T such that there does *not* exist a color X giving

$$T=\sigma S+(1-\sigma)X,$$

where σ is a positive fraction, while *every* color of the suite based on A, B can be represented by

$$\sigma S+\tau T, \quad \text{where } \sigma, \tau \geqslant 0, \sigma+\tau=1.$$

The whole class of boundary colors of unit brightness, such as S, T, will obviously form some closed one-dimensional manifold of colors, *the boundary* of the entire twofold of equibright colors. Notice, however, that we are by no means

entitled to say that this boundary consists, rigorously, of three or any finite number of color suites, although it is generally known that a "color triangle," i.e., a system of three color suites, based on a certain deep red (carmine), violet, and green, covers a large portion of the whole domain, nay, almost exhausts it. At any rate, it will be best at this stage not to prejudge at all the form of this boundary, but to think of it as of some closed set of colors, figuratively, a *closed color curve*.

Let A, B, C be three distinct colors of the twofold, within or on the boundary, but not belonging to one suite, i.e., not linearly dependent. Then

$$P=\xi A+\eta B+\zeta C, \quad \text{with} \quad \xi+\eta+\zeta=1 \qquad (2)$$

or

$$P=(uA+vB+wC)/(u+v+w) \qquad (2a)$$

will represent a two-dimensional manifold of colors, at any rate contained within, although not necessarily exhausting, the whole contemplated color twofold. To each set of non-negative ξ, η, ζ- or u, v, w-values corresponds a color of our twofold. *Vice versa*, every color of this submanifold (ABC) can be thus represented, *viz.* by three *homogeneous coordinates*, as ξ, η, ζ may be called.

So far the concept of a *distance* or a numerically expressible *difference* (distinctness) between any two colors, say

$$P=\xi A+\eta B+\zeta C \quad \text{and} \quad P'=\xi' A+\eta' B+\zeta' C,$$

is foreign to our circle of ideas. In other words, our color twofold has, so far, no metrical properties or, as the geometers say, is "amorphous," non-metrical, and will remain so until some system of metrics has been impressed upon it.

We may now turn to the question whether this color twofold (of constant brightness) is Desarguesian. The importance of this question, as has already been mentioned, lies therein that the validity of the celebrated *theorem of Desargues* is a necessary criterion for a projective threefold and also for a projective twofold, if the latter may be considered as embedded in, or forming a sub-domain of, the former. This theorem, taken geometrically, can be expressed as follows:

If two triads of points, A, B, C and A', B', C' are perspective from a point O, i.e., if the lines AA', BB', CC' are *copunctal* in O, then the three crosses of pairs of corresponding lines, $AB\cdot A'B'$, $BC\cdot B'C'$, and $CA\cdot C'A'$, are *collinear*, and conversely.

The line l common to these three crosses is called the axis of perspectivity, and the theorem can be stated more briefly, thus: *Two triangles perspective from a point* (O) *are perspective from a line* (l), *and conversely.*

Our question now is: Does the theorem of Desargues hold in the twofold of equibright colors, if "point" is replaced by "color" and "line" (i.e., straight line) by some one-dimensional *set* of colors?

The answer to this question will depend, of course, most essentially on the choice of the "set" of colors which is to correspond to a straight line. We have hitherto defined only one such set, *viz.* a color *suite*, a wholly non-metrical concept. A little later on, after having impressed some metrics upon the color twofold, we shall consider another type of a one-dimensional set, namely, the *shortest* color line or the color *geodesic*.

In the case of geometry the "lines," i.e., straight lines are at the same time the geodesics of the three-space or two-space (plane). Such, however, need not be the case in the manifold of colors. That is to say, the color suites need not (and, as we will presently have good reasons to believe, do not) coincide with the color geodesics of our twofold.

Let us, in the first place, substitute for the geometrical lines the *color suites*. Then the question as to the color-analog of the theorem of Desargues will run thus:

If two triads of colors, A, B, C and A', B', C' (neither triad on a suite), are perspective from a color O, are they also perspective from a color suite l, and conversely?

More explicitly, and in terms of actual experimental color mixing, the question and its clinching amounts to the following laboratory procedure:

Pick out some color O, and three more colors A, B, C not forming a suite. Blend O with A getting the color.

$$\left.\begin{aligned} A'&=\alpha A+(1-\alpha)O.\\ \text{Similarly,}\quad B'&=\beta B+(1-\beta)O,\\ C'&=\gamma C+(1-\gamma)O.\end{aligned}\right\}. \qquad (3)$$

Next, construct the color suites BC and $B'C'$, extending them (if necessary) sufficiently to find the color L common to these two suites, briefly, $L \equiv BC \cdot B'C'$. In much the same way determine the colors $M \equiv CA \cdot C'A'$ and $N \equiv AB \cdot A'B'$. Finally, test, always experimentally, whether the colors L, M, N do belong to the same suite, that is to say, whether a linear relation

$$\rho L + \sigma M + \tau N = 0, \quad \rho + \sigma + \tau = 0, [4] \qquad (4)$$

holds or not.

There seems to be no essential difficulty in actually performing such an experiment in the laboratory, selecting now one and now another septet of colors as O, A, B, C, and A', B', C' and in this manner exploring the whole twofold of equibright colors as to its Desarguesianism.

Such experimental test would, at any rate, be interesting and worth undertaking.

As a matter of fact, however, an affirmative answer to our question may be theoretically predicted if some elementary properties of color mixing are granted, namely, if the current view that the brightness of the resultant is always equal to the sum of the brightnesses of the components can be considered as firmly established, and that the sum or the difference of two correct color equations is again a valid color equation. The commutative and the associative properties of a color sum, manifestly indispensable for the legitimacy of the last statement, can certainly be considered as well established.

In fact, if these properties of color mixing be taken for granted, and if O, A, B, C are at unit brightness so also will be A', B', C' as expressed by (3); in fine, all these seven colors will belong to our twofold (brightness 1) of colors. Now, such being the case, the multiplication of the second of Eqs. (3) by $(1-\gamma)$ and of the third by $(1-\beta)$ and subtraction, side by side, will give

$$(1-\gamma)B' - (1-\beta)C' = (1-\gamma)\beta B - (1-\beta)\gamma C,$$

and since the right hand of this equation represents a color of the suite BC, and the left one of the suite $B'C'$, each of them represents, apart from a numerical factor, the color L common to BC and $B'C'$, to wit,

$$L = \frac{(1-\gamma)B' - (1-\beta)C'}{(1-\gamma)-(1-\beta)} = \frac{(1-\gamma)B' - (1-\beta)C'}{\beta-\gamma}$$

or $$(\beta-\gamma)L = (1-\gamma)B' - (1-\beta)C'.$$

Similarly, by cyclic permutation,

$$(\gamma-\alpha)M = (1-\alpha)C' - (1-\gamma)A',$$
$$(\alpha-\beta)N = (1-\beta)A' - (1-\alpha)B',$$

whence

$$(1-\alpha)(\beta-\gamma)L + (1-\beta)(\gamma-\alpha)M + (1-\gamma)(\alpha-\beta)N = 0,$$

say $$\rho L + \sigma M + \tau N = 0,$$

and here

$$\rho + \sigma + \tau = (1-\alpha)(\beta-\gamma) + (1-\beta)(\gamma-\alpha) + (1-\gamma)(\alpha-\beta),$$

i.e., $\rho+\sigma+\tau=0$, identically, as in (4). Thus L, M, N are colors at unit brightness *belonging to one suite*. Q. E. D.

This proof, based on the validity of the aforesaid properties of color mixing (which, if need be, can still be subjected to a more stringent experimental verification), establishes the color twofold of constant brightness as a *projective twofold*, comparable or, as the modern geometer would say, *isomorphic* with the Euclidean plane or, more generally, with the plane in any three-space of isotropic and constant curvature (*viz.* elliptic, parabolic or hyperbolic), provided that the role of "points" and "lines" is ascribed to colors and *color suites*.

Now, what is most important for an intrinsic (as contrasted with arbitrary, casual) numerical labeling of the colors of the contemplated twofold is that the validity of Desargues' theorem implies a powerful property, *viz.* the *uniqueness* of a *fourth harmonic color* to any given triad of colors A, B, C forming a suite. It is well known from projective geometry, that if Desargues' theorem holds, the construction of the fourth harmonic D to three collinear points A, B, C, conjugate to one of them, say A, is unique, that is to say, independent of the particular choice of the auxiliary lines which must be drawn in order to find D. It may be well to recall here the construction, in the convenient form due to von Staudt, of D, the fourth harmonic to A, B, C conjugate to A:

[4] Since L, M, N are all at unit intensity, we have, say $L = \xi M + (1-\xi)N$ or $-L + \xi M + (1-\xi)N = 0$, so that the sum of the three coefficients is $-1+\xi+1-\xi=0$.

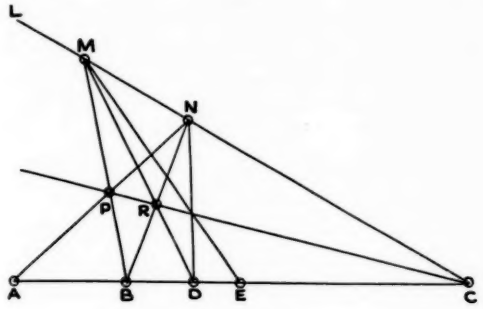


FIG. 1.

Through C draw any (straight) line l and take on it, arbitrarily, two points, M, N (Fig. 1). Draw the join of C and the cross $P \equiv AN \cdot BM$. Next, mark the cross $R \equiv PC \cdot NB$. Then the line MR will cross the base line ABC in the required fourth harmonic D, conjugate to A. And the remarkable property is that the position of D is wholly independent of the choice of the auxiliary points M and N, provided they are collinear with C. If any other pair M', N', always collinear with C, is chosen, then, as can readily be proved by Desargues' theorem, the result (D) is always the same. Briefly, A, B, C codetermine uniquely a fourth point, D, the harmonic conjugate to A.

Thus, replacing "point" by "color" and "line" by "color suite," we can find the *fourth harmonic color* (D) to any given suite of three colors (A, B, C), conjugate to one of these, say to A, unless D would fall beyond the actual boundary of the color twofold. This theoretical clause, however, drops out, if B is "between" A and C, that is to say, if B is obtainable by blending A with C. For then D, the harmonic conjugate to A, will always lie between, or be a mixture of, B and C, and will, therefore, be an actual color of the twofold.

Briefly, three colors A, B, C belonging to a suite will always co-determine a unique color of the same suite,

$$D = \text{harm. conjug. to } A.$$

Similarly, by a process indicated in Fig. 1, a unique fifth color of the same suite can be found, the fourth harmonic to B, D, C, say

$$E = \text{harm. conjug. to } B,$$

and so on. In fine, an unlimited number or, as we may say, a *Staudtian scale*, of colors, D, E, etc., all on the same suite (AC), can be constructed experimentally.

Imitating the well-known geometrical procedure, we may now attach definite numerical labels to all these colors, *viz.*, if the color A is given the label 0 and B the label 1, the colors D, E, etc., arrived at by 1, 2, etc., Staudtian constructions, will receive the labels 2, 3, and so on, and the color C which, theoretically, cannot be reached by a finite number of repetitions of the Staudtian constructiom, though it is more and more approached, will consistently be given the index ∞. To negative labels, -1, -2, etc. will correspond colors beyond one or the other of the basal colors A, C. Fractional (rational) labels, as $\frac{1}{2}$, $\frac{1}{3}$, $\frac{2}{3}$, $\frac{3}{2}$, etc., can be inserted by imitating the well-known processes of projective geometry. Any color S contained between two other colors R, T of the same suite will in this manner have for its label a number contained between the numbers serving as labels for R and T.

The labeling of A, B by 0 and 1 (and C by ∞) is, of course, an arbitrary, conventional act. But, this being decided upon, all the remaining colors of the suite AB are numbered (labeled) in a perfectly definite way, free from all arbitrariness. This process seems, therefore, to be a valuable means of constructing a rational colorimetric labeling.

Given any suite of colors l, the choice of A, C, or 0, ∞ and of B or 1 is, as has just been said, quite free. But it seems natural and particularly convenient to choose as the terminals A, C (0, ∞) the two boundary colors P and P' (Fig. 2) of the suite or, figuratively speaking, the crosses of the suite l with the boundary of the color twofold. These definite colors P, P', then, will be labeled with the Staudtian indices 0 and ∞. The color U to be labeled with the index 1 may still be picked out on l quite arbitrarily. But, if this be once chosen, the whole numerical scale of Staudtian indices

$$n = 0, 1, 2, 3, \cdots \infty$$

with all intermediate fractions, which are again uniquely determined, will be fixed. Since the two colors labeled 0 and ∞ are *boundary* colors,

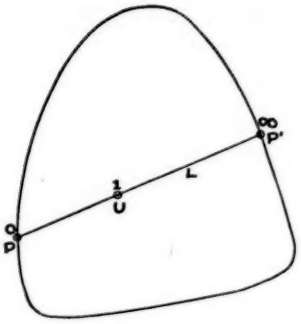


FIG. 2.

the whole suite l of colors will be covered by *positive* indices; colors which would correspond to negative indices n are nonexistent.

Furthermore, since every suite of colors has two definite terminals (as P, P') on the boundary of the twofold, the whole twofold of equibright colors will thus be intrinsically labeled by numbers, the Staudtian indices. This would seem a most rational method of labeling such a domain of colors.

A suitable scheme from the practical point of view would be to select some three boundary colors, say L, M, N, which, by blending, are apt to cover the largest portion of the whole color twofold, *viz.* a certain red, green, and violet, and to choose some color (not a boundary color) as our "unit" color U, i.e. so that U will receive the Staudtian index 1 on *each suite* of the twofold which contains U. For, manifestly, the whole twofold can be covered by drawing a pencil of suites through U (Fig. 3).

The triad L, M, N of "fundamental" colors, and the "unit" color U being chosen, every color of the twofold will be uniquely labeled by a pair of Staudtian indices or, as we will say, *projective coordinates* ξ and η, with (say) LM as the axis of ξ and LN as that of η, namely in the following way (Fig. 4).

Construct the color suite NU and protract it up to its intersection A with, or boundary color on, the suite LM. This color A will have the projective coordinates

$$\xi=1, \quad \eta=0, \qquad (A)$$

while L, M themselves will be labeled with

$$\xi=0, \quad \eta=0 \qquad (L)$$

and

$$\xi=\infty, \quad \eta=0. \qquad (M)$$

Similarly, the suite MU protracted will find on the suite LN a definite color B which will have the coordinates

$$\xi=0, \quad \eta=1, \qquad (B)$$

while L will be labeled as above and N will have the coordinates

$$\xi=0, \quad \eta=\infty, \qquad (N)$$

and the color U will receive a pair of units as its projective coordinates,

$$\xi=1, \quad \eta=1. \qquad (U)$$

Let, now, C be any color of the twofold bounded always by LMN. Then the suite NC, protracted, will find on the suite LM the projective coordinate ξ and MC protracted will find on the suite LN the coordinate η of C. Thus the projective coordinates ξ, η of any color C of the twofold will be uniquely determined.

In this scheme, the colors L, M, N themselves will have, as we just saw, the projective coordinates

$$\xi, \eta = 0, 0; \quad \infty, 0; \quad 0, \infty,$$

respectively. Further, the equation of the color suite LM will be just $\eta=0$, that of LN, $\xi=0$, while every color of the suite MN will have the projective coordinates $\xi=\infty$ and $\eta=\infty$.

This inconvenience and, at the same time, the privileged role of one of the three fundamental colors, *viz.* L, can be removed by a device familiar to geometers, namely, by introducing, instead of two, three so-called *homogeneous* projective coordinates, say x, y, z, to be defined by

$$\xi=x/z, \quad \eta=y/z.$$

Since in this manner the *ratios* only of x, y, z are fixed,

$$\lambda x, \lambda y, \lambda z$$

with any nonvanishing λ, will represent the same color of the twofold as

$$x, y, z.$$

The homogeneous projective coordinates of the colors L, M, N may now be taken to be

$$1, 0, 0; \quad 0, 1, 0; \quad 0, 0, 1,$$

those of the "unit" color U

$$1, 1, 1,$$

and the equations of the color suites MN, NL, and LM

$$x=0, \quad y=0, \quad z=0$$

respectively. Finally, any color suite of our twofold of equibright colors will be represented by a *homogeneous* linear equation in x, y, z, say

$$ax+by+cz=0$$

with constant coefficients a, b, c.

Let us now compare the homogeneous projective coordinates x, y, z of a color C of the twofold with the brightnesses u, v, w (say) of the three fundamental colors L, M, N, by the blending of which C is obtained. In fine, let

$$C=uL+vM+wN, \tag{5}$$

and our question is: How are x, y, z related to u, v, w?

The nonhomogeneous coordinates of C are $\xi=x/z$, $\eta=y/z$, and each of these was originally defined as the Staudtian index of A and B, the projections of C from N upon LM and from M upon LN (Fig. 4). Now, it can readily be shown (by returning to the Staudtian scale construction, Fig. 1) that if

$$L+M$$

is taken as the *unit* of the Staudtian scale on LM (index 1), the color A of the suite LM having the index ξ is, apart from an arbitrary factor,

$$L+\xi M,$$

and if it is to belong to our twofold (of brightness 1),

$$A=(L+\xi M)/(1+\xi).$$

Similarly, $\quad B=(L+\eta N)/(1+\eta).$

Now, C being the color common to the suites NA and MB, we have

$$C=\alpha A+(1-\alpha)N=\beta B+(1-\beta)M,$$

where α, β are some positive fractions. Thus, substituting the values of A and B, we have the equation

$$\frac{\alpha(L+\xi M)}{1+\xi}+(1-\alpha)N=\frac{\beta(L+\eta N)}{1+\eta}+(1-\beta)M,$$

and since L, M, N are linearly independent (i.e. do not belong to one suite), we have separately, equating the coefficients of L, M, N on both sides,

$$\frac{\alpha}{1+\xi}=\frac{\beta}{1+\eta}, \quad \frac{\alpha\xi}{1+\xi}=1-\beta, \quad 1-\alpha=\frac{\beta\eta}{1+\eta},$$

whence $\alpha=\dfrac{1+\xi}{1+\xi+\eta}, \quad 1-\alpha=\dfrac{\eta}{1+\xi+\eta}$

and $$C=\frac{(1+\xi)A+\eta N}{1+\xi+\eta}=\frac{L+\xi M+\eta N}{1+\xi+\eta},$$

and substituting $\xi=x/z$, $\eta=y/z$,

$$C=\frac{zL+xM+yN}{x+y+z}.$$

Thus, returning to (5), the required relations between the blending component brightnesses u,

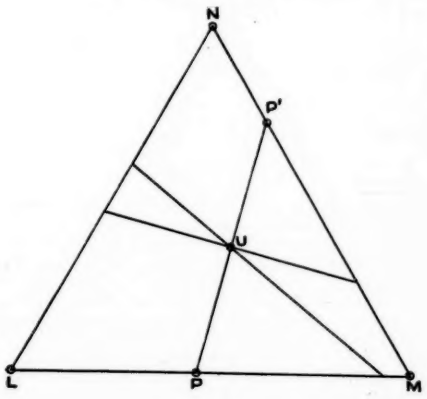


FIG. 3.

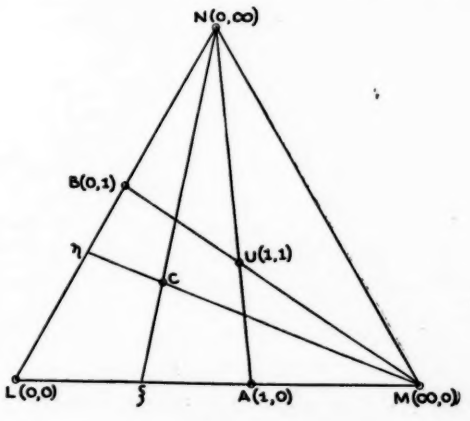


FIG. 4.

v, w and the homogeneous projective coordinates x, y, z of any color C of the twofold (of brightness 1) are found to be

$$u, v, w = (z, x, y)/(x+y+z). \qquad (6)$$

The relation $u+v+w=1$ is herewith satisfied automatically.

In words, *the component brightnesses of any color of our twofold are proportional to its homogeneous projective coordinates* (to wit, the first, second, and third component brightnesses are proportional to the third, first, and second homogeneous projective coordinates).

Since the ratios only of x, y, z are determined, one may as well put

$$x+y+z=1,$$

so that,
$$u, v, w = z, x, y. \qquad (6a)$$

The projective, and therefore intrinsic, coordinates, x, y, z of any color C can thus be found by comparatively simple color-blending operations,—for such operations only are required for determining u, v, w, while x, y, z as originally defined would call for very elaborate processes corresponding to Staudtian constructions. In virtue of the equalities (6a), of the two sets of coordinates the color equation (5) can now be written $C=zL+xM+yN$, or—if we rename our previous colors L, M, N by N, L, M, respectively,

$$C=xL+yM+zN. \qquad (7)$$

Before closing this, non-metrical, part of the investigation it will be well to emphasize once more that the Desargues theorem, with the uniqueness of the fourth harmonic and all other consequences, holds for the color twofold, *provided* that points and lines are replaced by colors and *color suites*, not color-geodesics (as will be seen presently, in the second part). Consequently, the representation of colors in the Euclidean plane is a valid representation of color-mixture phenomena, although, as shown later, it may not be a valid representation of the distinctness of color differences.

III. Metrics of the Color Domain (Twofold)

So far all our considerations concerning the twofold of equibright colors have been of an entirely non-metrical character. In other words, that color twofold has been non-metrical or amorphous. We now proceed to impress *metrics* upon it.

As before, let x, y, z, subject to the condition $x+y+z=1$, be the coordinates of a color C and $x+dx$, $y+dy$, $z+dz$,[5] those of a neighbor color C'.

What is the quantitative, numerical measure of the "difference" of these two colors or, in Helmholtz' terminology, the distinctness (Deutlichkeit) of their difference?

Needless to say, there is no such definite "measure," to begin with or given *a priori*: One may say only that a certain green, e.g., differs from a certain purple considerably more than some two green "shades" differ from each other, but not that the former difference is five or ten times the latter.

In order to be able to treat numerically the difference of any two colors whatever, we must first set up some differential expression for *the elementary difference* ds of a general pair of contiguous colors x, y, z and $x+dx$, $y+dy$, $z+dz$. By analogy with differential geometry, as initiated by Riemann, we may take for ds the square root of a *quadratic form* in dx, dy, dz, say an orthogonal form,[6]

$$ds^2=\gamma_1 dx^2+\gamma_2 dy^2+\gamma_3 dz^2, \qquad (8)$$

where each of the coefficients γ is some function of, in general, all three coordinates x, y, z. Of these three functions γ_1, etc. we may at this stage say only that they are continuous and positive, or nil at certain exceptional points (colors), but never negative.

Since we are dealing here with a twofold of colors, *viz.* of equal brightness (taken as unity), the coordinates are tied up by the relation $x+y+z=1$, and the differentials in (8) are subject to the condition

$$dx+dy+dz=0.$$

[5] We assume that the coordinates x, y, z of a color or the equivalent brightnesses of its components as perceived by the eye are liable of *continuous* variations, although it is undeniable that the eye distinguishes only a finite, discrete set of brightnesses between any pair of observable brightnesses x and $x+\Delta x$ (and similarly for y and z). This assumption is made only as a compromise, of course, so as to enable us to apply the rules of the differential calculus.

[6] One might consider the most general definite (i.e., always positive) quadratic form, containing also the rectangular terms $dxdy$, $dydz$, $dzdx$, but such a complication of treatment does not seem necessary at the present stage.

We may, therefore, as well eliminate one of the differentials, say dz, and use x, y as our independent variables. The difference element or, as we will sometimes say, the distance element (8) in the twofold of equibright colors will then become a binary quadratic form

$$ds^2=\gamma_1 dx^2+\gamma_2 dy^2+\gamma_3(dx+dy)^2$$

or
$$ds^2=g_{11}dx^2+2g_{12}dxdy+g_{22}dy^2, \tag{9}$$

where $g_{11}=\gamma_1+\gamma_3$, $g_{12}=\gamma_3$, $g_{22}=\gamma_2+\gamma_3$. (10)

Before trying to specify the form of g_{11}, g_{12}, g_{22} as functions of x, y ($z=1-x-y$), we may make a few general remarks of a metrical character.

The difference or, figuratively, the *distance* between any two colors of the twofold, $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ is defined by the value of the integral

$$s=\int_{P_1}^{P_2}(g_{11}dx^2+2g_{12}dxdy+g_{22}dy^2)^{\frac{1}{2}} \tag{11}$$

extended along a color curve which makes this integral a minimum or along a *geodesic* (shortest line) of the twofold.

A geodesic is defined by the variational equation

$$\delta\int_{P_1}^{P_2}ds=0, \tag{12}$$

the end-colors P_1, P_2 themselves being kept fixed in the variation process.

The differential equations of a color geodesic can be most easily derived from (12) as follows.

Let u be any auxiliary independent variable (integration variable) and let derivatives with respect to u be denoted by dots. Then (12) becomes

$$\delta\int_{u_1}^{u_2}\dot{s}du=0, \tag{12a}$$

where, by (9),

$$\dot{s}^2=g_{11}\dot{x}^2+2g_{12}\dot{x}\dot{y}+g_{22}\dot{y}^2. \tag{9a}$$

By the rules of the calculus of variations, (12a) gives the two differential equations

$$\frac{d}{du}\left(\frac{\partial\dot{s}}{\partial\dot{x}}\right)-\frac{\partial\dot{s}}{\partial x}=0, \quad \frac{d}{du}\left(\frac{\partial\dot{s}}{\partial\dot{y}}\right)-\frac{\partial\dot{s}}{\partial y}=0,$$

wherein one has to substitute, by (9a),

$$\dot{s}\frac{\partial\dot{s}}{\partial\dot{x}}=g_{11}\dot{x}+g_{12}\dot{y},$$

$$2\dot{s}\frac{\partial\dot{s}}{\partial x}=\dot{x}^2\frac{\partial g_{11}}{\partial x}+2\dot{x}\dot{y}\frac{\partial g_{12}}{\partial x}+\dot{y}^2\frac{\partial g_{22}}{\partial x}$$

and similar expressions for the derivatives with respect to $\dot{y}$ and y. Now, take for u the arc s of the geodesic itself. Then $\dot{s}=1$ and the required equations to a color geodesic of the twofold ultimately become

$$\begin{aligned}\frac{d}{ds}(g_{11}\dot{x}+g_{12}\dot{y})&=\tfrac{1}{2}\dot{x}^2\frac{\partial g_{11}}{\partial x}+\dot{x}\dot{y}\frac{\partial g_{12}}{\partial x}+\tfrac{1}{2}\dot{y}^2\frac{\partial g_{22}}{\partial x},\\ \frac{d}{ds}(g_{12}\dot{x}+g_{22}\dot{y})&=\tfrac{1}{2}\dot{x}^2\frac{\partial g_{11}}{\partial y}+\dot{x}\dot{y}\frac{\partial g_{12}}{\partial y}+\tfrac{1}{2}\dot{y}^2\frac{\partial g_{22}}{\partial y},\end{aligned} \tag{13}$$

where $\dot{x}=dx/ds$, etc. It can be easily shown that one of these equations is a consequence of the other, namely, in virtue of the obvious identical relation

$$g_{11}\dot{x}^2+2g_{12}\dot{x}\dot{y}+g_{22}\dot{y}^2=1. \tag{14}$$

The differential equations (13) being of the second order, a geodesic is fully determined by a color, say x_0, y_0, belonging to it and the values of $\dot{x}_0$, $\dot{y}_0$ or also by any two colors belonging to it.

If the system (13) of equations is solved for $\ddot{x}=d^2x/ds^2$ and $\ddot{y}$, one finds, after some reductions which may be omitted here,[7] the standard form of the geodesic equations

$$\frac{d^2x}{ds^2}+\left\{\begin{matrix}\alpha\beta\\1\end{matrix}\right\}\dot{x}_\alpha\dot{x}_\beta=0, \quad \frac{d^2y}{ds^2}+\left\{\begin{matrix}\alpha\beta\\2\end{matrix}\right\}\dot{x}_\alpha\dot{x}_\beta=0, \tag{13a}$$

where x_1, x_2 stand for x, y and $\left\{\begin{matrix}\alpha\beta\\\kappa\end{matrix}\right\}$ are the well-known Christoffel symbols, to be summed over all values of α, β, in our case 1, 2.

If the coefficients g_{11}, etc. are all constant, the Christoffel symbols vanish and we are left with

$$d^2x/ds^2=0, \quad d^2y/ds^2=0,$$

so that x and y are linear functions of s with

[7] It will suffice to mention that $\ddot{x}$ and $\ddot{y}$ occur only in the left-hand members of (13) which, in developed form, are

$$g_{11}\ddot{x}+g_{12}\ddot{y}+\dot{x}\left(\frac{\partial g_{11}}{\partial x}\dot{x}+\frac{\partial g_{11}}{\partial y}\dot{y}\right)+\dot{y}\left(\frac{\partial g_{12}}{\partial x}\dot{x}+\frac{\partial g_{12}}{\partial y}\dot{y}\right), \text{ etc.}$$

constant coefficients, and if s is eliminated between them, we have a linear relation between x and y. The same follows, more directly, from the original equations (13) which for constant g_{11}, etc. become $g_{11}\dot{x}+g_{12}\dot{y}=c_1$, $g_{12}\dot{x}+g_{22}\dot{y}=c_2$, whence $g_{11}x+g_{12}y=c_1 s+\text{const.}$, etc. In this case, then, the geodesics of our twofold would be color suites, i.e., continua of colors obtainable by simple blending of some two colors.

If, however, all or only some of the coefficients g_{11}, g_{12}, g_{22} are not constant, *the color suites do not coincide with the geodesics*, and then the distinctness properties of the color twofold cannot be correctly represented by a geometrical plane, and not even by a surface of constant curvature. That is to say, the geometrical distances between pairs of points of such a surface will, in general, not be proportional to the differences between the two colors represented by these points.

We will presently find reasons to believe that such (non-representability of color discrimination within the color twofold by a plane) is actually the case.

Least perceptible differences utilized

The coefficients g_{11}, g_{12}, g_{22} of the quadratic form (9) representing the distance of two colors, which so far were treated as entirely unknown, must be determined experimentally in some way or other. Now, it seems most natural to base this determination upon a characteristic feature of the color domain, namely, the existence of *least visually perceptible differences* with any given color P as starting point.

Let $P(x, y)$ be some color at unit brightness. From P we can pass to other, neighbor colors by mixing P with a small amount of any of the boundary colors, say Q. In the suite of colors PQ, all at unit brightness, there will be a color R *just discernible* from P itself, say,

$$R=\tilde{\omega}Q+(1-\tilde{\omega})P,$$

where $\tilde{\omega}$ is some fraction, presumably small, to be evaluated experimentally. The difference or distance between P and R may conveniently be taken as *unit distance* along PQ,

$$\text{distance } [P, R]=1.$$

In this manner, taking in succession different boundary colors Q, the given color P can be surrounded, figuratively speaking, by a closed sequence of colors R, all at unit distance from P, as shown symbolically on Fig. 5. The totality of these colors R may properly be called a *unit circle* about P as center. The equation of this "circle" of colors is, by its very definition,

$$\text{distance } [P, \tilde{\omega}Q+(1-\tilde{\omega})P]=1, \qquad (15)$$

where Q is any boundary color and $\tilde{\omega}$ a numerical, fractional coefficient which will, in general, vary while Q moves along the boundary of the twofold. On the other hand, the equation of such a unit circle in terms of the quadratic differential form (8), with x, $y(z=1-x-y)$ as the coordinates of P and $x+dx$, $y+dy$ as those of R,[8] will be

$$\gamma_1 dx^2+\gamma_2 dy^2+\gamma_3 dz^2=1, \qquad (16)$$

where $\gamma_1=\gamma_1(x, y)$, etc., belong to P. (Pedantically, we might be expected to set in the last two equations, instead of 1, a constant small fraction, say $1/n^2$, where n is a large number. But, since this constant n^2 may be thrown as a factor upon all three γ's, we may as well omit it.)

Now, by comparing (16) with (15), where $\tilde{\omega}$ is to be evaluated experimentally, the form of γ_1, etc. and thence that of g_{11}, etc. as functions of x, y can be determined. A *color geodesic* joining any pair of colors A, B will now be that chain of colors leading from A to B which consists of the least number of just perceptible differences or, as we may say, unit steps. As has already been mentioned, such a geodesic need by no means coincide, in general, with the color suite AB.

Provided, then, that $\tilde{\omega}$ has been determined experimentally, the coefficients g_{11}, etc. can now be readily determined by comparing (15) with (16). In fact, let, as before, x, y, and $z=1-x-y$ be the coordinates of P and x', y', $z'=1-x'-y'$ those of Q,

$$P=xL+yM+zN, \quad Q=x'L+y'M+z'N.$$

Then

$$R=(1-\tilde{\omega})P+\tilde{\omega}Q=[(1-\tilde{\omega})x+\tilde{\omega}x']L+\text{etc.}$$
$$=(x+dx)L+\text{etc.};$$

whence $dx=\tilde{\omega}(x'-x)$, $dy=\tilde{\omega}(y'-y)$,

[8] To consider the differences of the coordinates of P and R as infinitesimal (dx, dy) is an approximately legitimate act provided that $\tilde{\omega}$ is actually a small fraction of unity.

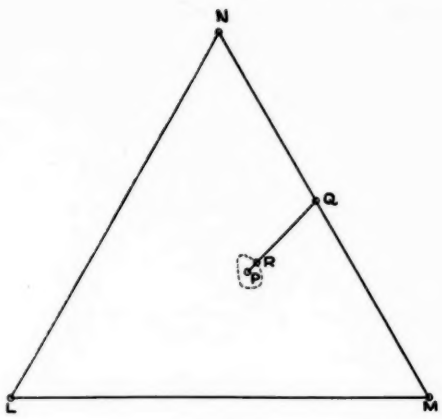


FIG. 5.

and $dz=\tilde{\omega}(z'-z)=-dx-dy$, automatically, and, substituting these values of dx, dy into (16),

$$\tilde{\omega}^2\{\gamma_1(x'-x)^2+\gamma_2(y'-y)^2+\gamma_3(z'-z)^2\}=1. \quad (17)$$

This holds for any $Q(x', y')$ picked out on the boundary of the twofold.

Now, let Q coincide with L, one of our three fundamental colors, and let the corresponding $\tilde{\omega}$ be denoted by $\tilde{\omega}_1$. Then, since $x'=1$, $y'=z'=0$, formula (17) will become

$$\gamma_1(x-1)^2+\gamma_2y^2+\gamma_3z^2=1/\tilde{\omega}_1^2.$$

Two similar equations can be written down for the case when Q coincides with M or N. Denoting the corresponding values of $\tilde{\omega}$ by $\tilde{\omega}_2$ and $\tilde{\omega}_3$, we have in all three linear equations,

$$\begin{aligned}(1-x)^2\gamma_1+y^2\gamma_2+z^2\gamma_3&=1/\tilde{\omega}_1^2,\\ x^2\gamma_1+(1-y)^2\gamma_2+z^2\gamma_3&=1/\tilde{\omega}_2^2,\\ x^2\gamma_1+y^2\gamma_2+(1-z)^2\gamma_3&=1/\tilde{\omega}_3^2,\end{aligned} \quad (18)$$

for the determination of γ_1 γ_2, γ_3 in terms of the explicit coordinates x, y, z and of $\tilde{\omega}_1$, $\tilde{\omega}_2$, $\tilde{\omega}_3$ which are some functions of the coordinates to be empirically explored.

It remains to solve this system for γ_1, etc. Now, the determinant of (18) is

$$\begin{aligned}\mathfrak{F}=&(1-x)^2(1-y)^2(1-z)^2-(1-x)^2y^2z^2\\ &-(1-y)^2z^2x^2-(1-z)^2x^2y^2+2x^2y^2z^2\end{aligned}$$

or, after easy reductions, and taking account of the condition $x+y+z=1$,

$$\mathfrak{F}=2xyz.$$

Thus,

$$\begin{aligned}2xyz\cdot\gamma_1&=\frac{x}{\tilde{\omega}_1^2}(x+2yz)-\frac{y^2}{\tilde{\omega}_2^2}(1-2z)-\frac{z^2}{\tilde{\omega}_3^2}(1-2y),\\ 2xyz\cdot\gamma_2&=\frac{y}{\tilde{\omega}_2^2}(y+2zx)-\frac{z^2}{\tilde{\omega}_3^2}(1-2x)-\frac{x^2}{\tilde{\omega}_1^2}(1-2z),\\ 2xyz\cdot\gamma_3&=\frac{z}{\tilde{\omega}_3^2}(z+2xy)-\frac{x^2}{\tilde{\omega}_1^2}(1-2y)-\frac{y^2}{\tilde{\omega}_2^2}(1-2x),\end{aligned} \quad (19)$$

whence, by (10)

$$\begin{aligned}g_{11}&=(1-y)\left(\frac{1}{\tilde{\omega}_1^2z}+\frac{1}{\tilde{\omega}_3^2x}\right)-\frac{y^2}{\tilde{\omega}_2^2zx},\\ g_{22}&=(1-x)\left(\frac{1}{\tilde{\omega}_2^2z}+\frac{1}{\tilde{\omega}_3^2y}\right)-\frac{x^2}{\tilde{\omega}_1^2yz},\\ g_{12}&=\frac{1}{2xyz}\left\{\frac{z}{\tilde{\omega}_3^2}(z+2xy)-\frac{x^2}{\tilde{\omega}_1^2}(1-2y)-\frac{y^2}{\tilde{\omega}_2^2}(1-2x)\right\}.\end{aligned} \quad (20)$$

These, with $x+y+z=1$, are the required coefficients of the distance element

$$ds^2=g_{11}dx^2+2g_{12}dxdy+g_{22}dy^2. \quad (9\ bis)$$

They contain x, y, z explicitly and also through $\tilde{\omega}_1$, etc. which are certainly not constant.

IV. Geometrical Properties of a Color Domain Obeying Fechner's Law

A. The twofold of colors

The safest way to evaluate g_{11}, g_{12}, g_{22} would be to determine the $\tilde{\omega}$'s experimentally in accordance with their definition given above. It seems, however, worth while at this stage (while extensive measurements of the $\tilde{\omega}$'s are not, to the writer's knowledge, available) to inquire what the coefficients γ_1, etc. and g_{11}, etc. become, if the $\tilde{\omega}$'s are assumed to obey the simple Fechner law.

The meaning of $\tilde{\omega}$ has been fixed by (15). We have, e.g., for $\tilde{\omega}_1$, when Q coincides with L,

$$\text{distance}\,[P,\ \tilde{\omega}_1L+(1-\tilde{\omega}_1)P]=1,$$

and this is the least perceptible step from the color $P(x, y, z)$ towards $L(x'=1, y'=z'=0)$. Now,

since x is the amount (brightness) of the color L already contained in P and $\bar{\omega}_1$ is the least perceptible addition to it, we would have, by Fechner's law,

$$\bar{\omega}_1 = \alpha_1 x,$$

where α_1 is a constant. [More accurately, $\bar{\omega}_1 = \alpha_1 x(1-\bar{\omega}_1)$ or, since $\bar{\omega}_1$ can be supposed to be a small fraction of unity, $\bar{\omega}_1 \doteqdot \alpha_1 x - \alpha_1^2 x^2$, but the last, second-order, term can be neglected.]

Similarly, $\quad \bar{\omega}_2 = \alpha_2 y, \quad \bar{\omega}_3 = \alpha_3 z,$

where α_2, α_3 are constants, not necessarily equal to α_1.

Thus, if Fechner's law be accepted, we have, by (19),

$$\gamma_1 = \frac{1}{2xyz}\left\{\frac{1}{\alpha_1^2}\left(1+\frac{2yz}{x}\right) - \frac{1-2z}{\alpha_2^2} - \frac{1-2y}{\alpha_3^2}\right\}, \text{ etc.,} \tag{21}$$

by cyclic permutation of x, y, z and 1, 2, 3, and, by (20),

$$g_{11} = \frac{1}{xz}\left\{(1-y)\left(\frac{1}{\alpha_1^2 x}+\frac{1}{\alpha_3^2 z}\right) - \frac{1}{\alpha_2^2}\right\},$$

$$g_{22} = \frac{1}{yz}\left\{(1-x)\left(\frac{1}{\alpha_2^2 y}+\frac{1}{\alpha_3^2 z}\right) - \frac{1}{\alpha_1^2}\right\}, \tag{22}$$

$$g_{12} = \gamma_3 = \frac{1}{2xyz}\left\{\frac{1}{\alpha_3^2}\left(1+\frac{2xy}{z}\right) - \frac{1-2y}{\alpha_1^2} - \frac{1-2x}{\alpha_2^2}\right\},$$

where $x+y+z=1$.

Now, these coefficients are neither constants nor reducible to constants by a linear transformation of the coordinates.

Consequently, by what we saw before, if the Fechner law holds, the geodesica of the color twofold of constant brightness do *not* coincide with the color suites.

Such will still be the case if it should turn out that the three Fechner law coefficients are all equal, say

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha.$$

Although such equality need by no means hold, we may limit our further discussion to this simplest case.

The formulae (21) are then reduced to

$$\gamma_1 = (x-2x^2+2yz)/(2\alpha^2 x^2 yz), \text{ etc.,}$$
$$\text{by cyclic permutation,} \tag{21a}$$

and the formulae (22) become

$$g_{11} = \frac{1}{\alpha^2 xz}\left[\frac{(1-y)^2}{xz} - 1\right],$$

$$g_{22} = \frac{1}{\alpha^2 yz}\left[\frac{(1-x)^2}{yz} - 1\right], \tag{22a}$$

$$g_{12} = \gamma_3 = \frac{z-2z^2+2xy}{2\alpha^2 xyz^2},$$

with $x+y+z=1$.

Before proceeding further it may be well to remark that each of the γ's vanishes along a whole color line or curve, *viz.* $\gamma_1=0$ along the line

$$(C_1) \quad x(x-\tfrac{1}{2}) = yz, \quad \text{with} \quad x+y+z=1,$$

and similarly $\gamma_2=0$, $\gamma_3=0$ along the lines

$$(C_2) \quad y(y-\tfrac{1}{2}) = zx, \qquad (C_3) \quad z(z-\tfrac{1}{2}) = xy.$$

This, however, does not give rise to any difficulty. In fact, if a color P is on the $\gamma_1=0$ line, we have for all elements starting from P

$$ds^2 = \gamma_2 dy^2 + \gamma_3 dz^2,$$

and if P is the cross of, say, the $\gamma_2=0$ and the $\gamma_3=0$ lines,

$$ds^2 = \gamma_1 dx^2.$$

In neither case does ds vanish. It can, moreover, be readily proved that there is no such color P at which ds vanishes altogether. In other words, the three lines under discussion have no common intersection. For, if they had one, the three equations (C_1), (C_2), (C_3) would hold simultaneously. Now, their product would give, in view of $x+y+z=1$,

$$yz+zx+xy = \tfrac{1}{4},$$

and their sum,

$$yz+zx+xy = x^2+y^2+z^2-\tfrac{1}{2},$$

whence $\qquad x^2+y^2+z^2 = \tfrac{3}{4}.$

On the other hand, since the system of equations (C_1), (C_2), (C_3) remains the same when x, y, z are interchanged, we must have $x=y=z=\frac{1}{3}$ and, therefore,

$$x^2+y^2+z^2 = \tfrac{1}{3},$$

which clashes with the value just found. Consequently, the three singular lines do not intersect in a point (color) and ds does never vanish. And since γ_1, γ_2, γ_3 do not vanish simultaneously, g_{11}, g_{12}, g_{22} cannot vanish simultaneously.

The distance element (9) in the twofold of equibright colors with the values (22a) of g_{11}, etc. now becomes

$$ds^2 = \frac{1}{\alpha^2 z}\left\{\frac{1}{x}\left[\frac{(1-y)^2}{xz} - 1\right]dx^2 + \frac{1}{y}\left[\frac{(1-x)^2}{yz} - 1\right]dy^2 + \frac{z-2z^2+2xy}{xyz}dxdy\right\}, \tag{23}$$

where $z=1-x-y$.

The geodesics of this twofold, which are given by (13), are manifestly not expressible by linear relations between x, y, z and do not, therefore, coincide with the color suites.

Thus, the geodesics generated by Fechner's law in the twofold of equibright colors do not coincide with the color suites of the twofold. This fact is, of course, entirely independent of any space-geometrical considerations. In fine, it is a property of the colors themselves and their mutual relations with regard to blending (suites) and numbers of least perceptible differences or steps (geodesics) defined by Fechner's law. No matter how our color twofold is represented graphically, geometrically, the lines representing color suites will deviate from the lines representing color geodesics.

If we now wish to construct in accordance with Fechner's law a *faithful* map of the distinctness properties of the color twofold on a geometrical surface, i.e. so that color geodesics should be represented by the geodesics of the surface and that differences between any two colors A, B should be proportional to the distance AB measured along the surface geodesics, with a constant coefficient of proportionality, we are not free to pick out for this purpose, at random, a plane or some other surface we fancy, but we must select a surface which has everywhere the same *Gaussian curvature* K as our color twofold. This "curvature" of a twofold, which is by no means dependent on any spacial imagery, is a certain *invariant*[9] differential expression containing the coefficients g_{11}, g_{12}, g_{22} of the quadratic form expressing the distance element ds^2 of that twofold together with their first and second derivatives, namely, if g be the determinant of the quadratic form, i.e. $g_{11}g_{22}-g_{12}^2$,

$$K=\frac{1}{2\sqrt{g}}\left(\frac{\partial A}{\partial x}+\frac{\partial B}{\partial y}\right),$$

$$\text{where}\quad A=\frac{1}{\sqrt{g}}\left(\frac{g}{g_{11}}\frac{\partial g_{11}}{\partial y}-\frac{\partial g_{22}}{\partial x}\right), \tag{24}$$

$$B=\frac{1}{\sqrt{g}}\left(2\frac{\partial g_{12}}{\partial x}-\frac{\partial g_{11}}{\partial y}-\frac{g_{12}}{g_{11}}\frac{\partial g_{11}}{\partial x}\right).$$

[9] Invariant with respect to any transformations whatever of the coordinates x, y.

The value of K at any point (color) x, y is independent of the choice of the system of coordinates and is, in this sense, an *intrinsic property* of the twofold. If K were zero everywhere, the color twofold could be faithfully represented on a Euclidean plane, and if K were a positive or a negative constant, it could be represented on a piece of a spherical or pseudospherical surface respectively, the latter being the well-known surface of revolution of a tractrix, resembling a champagne glass. But if K is not constant, and such indeed will be the case of the Fechnerian color twofold, we must use for its geometrical representation a more complicated surface, of variable curvature, namely one whose Gaussian curvature has at every point a value equal to that of K at the color (x, y) corresponding to that point.

In order to find K explicitly one has only to substitute in (24) the coefficients g_{11}, etc. of dx^2, $2dxdy$, and dy^2 as given by (22a) and rewritten in (23).

One finds, after some lengthy but straightforward reductions,

$$\frac{\partial g_{11}}{\partial y}=\frac{1+x-y}{\alpha^2xz^3},\quad \frac{\partial g_{22}}{\partial x}=\frac{1-x+y}{\alpha^2yz^3},$$

$$\frac{\partial g_{11}}{\partial x}=\frac{(z-x)[1-4(1-y)^2]}{\alpha^2x^2z^2},$$

$$\frac{\partial g_{12}}{\partial x}=\frac{xz(2zx-4z+1)+z^2(2z-1)+4x^2y}{2\alpha^2x^2yz^3},$$

$$\text{and}\quad A=\frac{(x+y)^2-2y^3+2x+6y-3}{2\alpha^2(g)^{1/2}\cdot yz^3[(1-y)^2-xz]} \tag{25}$$

$$\alpha^2x^2yz^3(g)^{1/2}\cdot B=xzT+z^2(2z-1)-xy(1-3x-y),$$

$$T=2xz-4z+1 \tag{26}$$

$$-\frac{z-x}{2}\cdot\frac{[1-4(1-y)^2][z-2z^2+2xy]}{(1-y)^2-xz},$$

where

$$g=\frac{1}{\alpha^4x^2y^2z^4}\{[(1-y)^2-xz][(1-x)^2-yz]$$

$$-\tfrac{1}{4}[z-2z^2+2xy]^2\}\equiv\frac{g_n}{\alpha^4x^2y^2z^4},\ \text{say.} \tag{27}$$

The expressions for A and B whose derivatives are required for the curvature K, are thus seen to be extremely complicated. It would, under these circumstances, be pretty useless to write out the complete formula for K in terms of x, y. What interests us here is to show that K is not constant, and for this purpose it is enough to evaluate it for some two different colors and to show that these values differ from each other.

Let A_n, B_n be the numerators of A, B, as given in (25) and (26), so that

$$A=\frac{A_n}{2\alpha^2yz^3[(1-y)^2-x^2]\sqrt{g}},\quad B=\frac{B_n}{\alpha^2x^2y^2z^3\sqrt{g}}.$$

Then, remembering that $z=1-x-y$,

$$2\alpha^2y\frac{\partial A}{\partial x}=\frac{\partial A_n/\partial x}{z^3[(1-y)^2-x^2]\sqrt{g}}$$
$$-\frac{A_n}{2z^6[(1-x)^2-x^2]^2y^3}\left\{z^3[(1-y)^2-x^2]\frac{\partial g}{\partial x}\right.$$
$$\left.-2gz^2[3([1-y]^2-x^2)+2xz]\right\},$$

$$\alpha^2x^2\frac{\partial B}{\partial y}=\frac{\partial B_n/\partial y}{yz^3\sqrt{g}}-\frac{B_n}{2y^2z^6g^{\frac{3}{2}}}\left[2z^2g(z-3y)+yz^3\frac{\partial g}{\partial y}\right],$$

where, by (25), (26),

$$\frac{\partial A_n}{\partial x}=2(x+y)+2,$$

$$\frac{\partial B_n}{\partial y}=xz\frac{\partial T}{\partial y}-xT+2z-6z^2-x(1-3x-2y),$$

$$\frac{\partial T}{\partial y}=4-2x+\frac{[1-4(1-y)^2][z-2z^2+2xy]}{2[(1-y)^2-xz]}$$
$$-\frac{z-x}{2}\frac{\partial}{\partial y}\left\{\frac{[1-4(1-y)^2][z-2z^2+2xy]}{(1-y)^2-xz}\right\},$$

and

$$\frac{\partial g}{\partial x}=\frac{xz\partial g_n/\partial x-2g_n(z-2x)}{\alpha^4y^2x^3z^5},$$

$$\frac{\partial g_n}{\partial x}=[xz-(1-y)^2](y+2z)+[yz-(1-x)^2](z-x)$$
$$-\tfrac{1}{2}(z-2z^2+2xy)(3-2y-4x),$$

$$\frac{\partial g}{\partial y}=\frac{yz\partial g_n/\partial y-2g_n(z-2y)}{\alpha^4x^2y^3z^5},$$

$$\frac{\partial g_n}{\partial y}=[yz-(1-x)^2](x+2z)+[xz-(1-y)^2](2-y)$$
$$-\tfrac{1}{2}(z-2z^2+2xy)(3-2x-4y).$$

Now, as we have just said, to ascertain that the curvature K is neither zero nor constant, it will be enough to evaluate K numerically, by (24) and all these auxiliary expressions, for a couple of colors, i.e., pairs of numerical values of x, y (and $z=1-x-y$), let us say,

$$K=K_a, \text{ for } x=y=z=\tfrac{1}{3},$$

and $$K=K_b, \text{ for } x=z=\tfrac{1}{4},\ y=\tfrac{1}{2}.$$

In the first case we have

$$\frac{\partial A_n}{\partial x}=\frac{10}{3},\quad \frac{\partial T}{y}=\frac{53}{18},\quad T=-\frac{1}{9},\quad \frac{\partial B_n}{\partial y}=\frac{95}{162},$$

$$A_n=\frac{1}{27},\quad B_n=-\frac{1}{81},\quad g_n=\frac{1}{12},\quad g=\frac{3^7}{4\alpha^4},$$

$$\frac{\partial g_n}{\partial x}=\frac{\partial g_n}{\partial y}=-\frac{1}{2},\quad \frac{\partial g}{\partial x}=\frac{\partial g}{\partial y}=-\frac{3^8}{\alpha^4},\quad \frac{\partial A}{\partial x}=\frac{85}{3\sqrt{3}},$$
$$\frac{\partial B}{\partial y}=\frac{35}{3\sqrt{3}},\quad \frac{\partial A}{\partial x}+\frac{\partial B}{\partial y}=\frac{40}{\sqrt{3}},$$

whence, finally,

$$K_a=(40/81)\alpha^2\doteqdot 0.4938\alpha^2.$$

In much the same way one finds, for the second case,

$$K_b=(1711/108)\alpha^2\doteqdot 15.84\alpha^2.$$

At both colors, then, the Gaussian curvature of the color twofold is positive, but at $x=z=\frac{1}{4}$, $y=\frac{1}{2}$ it is a little over 32 times greater than at $x=y=z=\frac{1}{3}$, which certainly is a very pronounced variation. This proves our statement.

The curvature K becomes *infinite* all along the periphery LMN of the contemplated color twofold, where either x or y or z vanish. It is enough to prove this for $z=0$, i.e. for the line LM, where $x+y=1$. (For MN and NL play exactly the same roles as LM.) Now, for z tending to zero the determinant tends to

$$g=-2/\alpha^4xy\cdot 1/z^3;$$

further, $\partial A_n/\partial x=4$ and, since $\partial T/\partial y$ remains finite, $\partial B_n/\partial y=-x[T+1-3x-2y]$,

$$T=1+(1-4x^2)y, \quad \partial B_n/\partial y=x[x-(1-4x^2)y],$$

so that $\partial A/\partial x$ and $\partial B/\partial y$ become infinite of the order $1/z^6$ at least, and K tends to infinity at least as $z^{3/2-6}=z^{-9/2}$.

The border LMN of our color twofold, or rather of the surface representing it, is thus a sharp ridge.

Our principal object here was to show that the Fechnerian color twofold can be faithfully represented only on a surface of variable curvature.

We have so far considered a single twofold of equibright colors, *viz.* $x+y+z=1$. Any other such twofold, of brightness i, can be covered by our preceding formulae, provided that $x+y+z=1$ is replaced by

$$x+y+z=i,$$

where i is a constant number. Our previous formulae can, therefore, be retained if only x, y, z are replaced by x/i, y/i, z/i. Thus, if we still write

$$ds^2=g_{11}dx^2+2g_{12}dxdy+g_{22}dy^2,$$

the coefficients of this quadratic form will become

$$g_{11}=\frac{1}{\alpha^2xz}\left[\frac{(i-y)^2}{xz}-1\right], \quad g_{12}=\frac{iz-2z^2+2xy}{2ixyz},$$

$$g_{22}=\frac{1}{\alpha^2yz}\left[\frac{(i-x)^2}{yz}-1\right],$$

whence the curvature and all other relevant properties of the color twofold of brightness i can be readily derived. The Fechner law coefficient α need not be an absolute constant; it may possibly depend on the brightness i. This question can be decided only by experiment. Such also is the position of the question whether the original coefficients α_1, α_2, α_3 are all equal or not. An extension of the above set of formulae to the case of different α's offers no difficulty.

B. The threefold of colors

We have so far treated a submanifold of the whole domain of colors, *viz.* a twofold of equibright colors, $x+y+z=$const. We will conclude the present paper by considering the main intrinsic properties of the complete domain of colors, which is a threefold.

The brightness coordinates, i.e. the brightnesses u, v, w of the components L, M, N of any color P, which can again be written

$$P=uL+vM+wN, \tag{28}$$

are now unrestricted, positive numbers. Thus the domain now to be considered is a *threefold*. The brightness of P is

$$i=u+v+w. \tag{29}$$

Let $P_1(u_1, v_1, w_1)$ and $P_2(u_2, v_2, w_2)$, of brightness i_1, i_2, be any two colors of this threefold. They can be mixed or blended in any proportions, giving the resultant

$$R=c_1P_1+c_2P_2$$

of brightness $i=c_1i_1+c_2i_2$, where c_1, c_2 are any positive numbers. Thus the blends R of P_1, P_2 fill out a whole *twofold* of colors representable by some surface. Our previous concept of a one-dimensional "suite of colors" now falls to pieces. For there is no reason to give to the brightness $i=c_1i_1+c_2i_2$ of R any one rather than any other value. In other words the color twofold R is cut by the twofolds of equibright colors in an infinity of color suites. In fine, two colors P_1, P_2 no longer determine a unique suite, and we can no longer avail ourselves of the single concept of a "color suite" in order to discuss the question of projectivity or desarguesianism of the threefold of colors. Consequently the concepts of a fourth harmonic color, of Staudtian indices, and of projective coordinates (like the previous ξ, η or x, y, z) are not applicable to the color threefold and, to begin with, we are bound to use as "coordinates" of a color P its component brightnesses u, v, w.

The simple color suite having lost its importance, what now is uniquely determined by any pair of colors P_1, P_2 of the threefold is the geodesic, that is to say, a one-dimensional set of colors which leads from P_1 to P_2 by the *least* (or stationary) number of just perceptible steps.

We will again suppose that the elementary difference or distance ds between two neighbor colors u, v, w and $u+du$, $v+dv$, $w+dw$ is given by an orthogonal quadratic form of du, dv, dw, say

$$ds^2=g_1du^2+g_2dv^2+g_3dw^2, \tag{30}$$

where g_1, etc. are some functions of, in general, all three coordinates u, v, w. The exact nature of these functions might be wholly determined by experiments. Yet, their form may to a certain extent be predicted by utilizing, in the first place, the existence of least perceptible differences in brightness with the Fechner law applied to brightness variations, while hue and saturation are kept fixed, and, in the second place, by considering elementary distances within the twofold of equibright colors passing through the point (color) P.

In fact, consider the one-dimensional set of colors which have the same so-called "dominant wave-length and purity" as the given color $P(u, v, w)$, i.e. all the colors

$$P' = \kappa P = \kappa(uL + vM + wN),$$

where u, v, w are kept constant, while κ is variable. This may be called an *isochrome* line. The brightness of P being i, that of any P' will be

$$i' = \kappa i = \kappa(u+v+w).$$

Now, let the color P' just distinguishable from P be that whose brightness is

$$i' = \kappa i = (1+\sigma)i, \tag{31}$$

where σ is some, supposedly small, fraction and let the difference or distance ds between P and this color be again taken as unity. Then, since in this case

$$du = u' - u = \sigma u, \quad dv = \sigma v, \quad dw = \sigma w,$$

we shall have, by (30),

$$g_1u^2 + g_2v^2 + g_3w^2 = 1/\sigma^2. \tag{32}$$

The numerical value of σ, which may depend on u, v, w, should be determined experimentally. If the simple Fechner law were valid for brightness, σ would be a constant. But it is better to leave the constancy or variability of σ an open question.

So far, then, we have a single equation, (32), for the three unknown functions g_1, etc. of u, v, w, in which, moreover, σ itself is to be viewed as a function of u, v, w to be found experimentally. In the next place, however, we can set up more equations for the g's by considering the twofold of equibright colors laid through P, for which, that is,

$$u+v+w = i = \text{const.},$$

and, therefore, $du+dv+dw=0$. For any color $u+du$, $v+dv$, $w+dw$ contained in this twofold we have, by (30), eliminating dw,

$$ds^2 = (g_1+g_3)du^2 + (g_2+g_3)dv^2 + 2g_3 du\,dv$$

or $$ds^2 = i^2\{(g_1+g_3)du'^2 + (g_2+g_3)dv'^2 + 2g_3 du'dv'\},$$

where $u' = u/i$, etc., so that $u'+v'+w' = 1$. Now, since we deal here with a twofold of equibright colors, $i^2(g_1+g_3)$, $i^2(g_2+g_3)$, and i^2g_3 will be the same functions of u', v', w' as were our previous $\gamma_1+\gamma_3$, $\gamma_2+\gamma_3$, and γ_3 of x, y, z, and therefore also i^2g_1, i^2g_2, i^2g_3 will be the same functions of u/i, v/i, w/i as were γ_1, γ_2, γ_3 of x, y, z, i.e., if we again limit ourselves to the case $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$, by (22a),

$$i^2g_1 = \frac{u/i - 2(u^2 - vw)/i^2}{2\alpha^2 u^2 vw/i^4}, \text{ etc.},$$

i.e.,

$$g_1 = \frac{iu - 2(u^2-vw)}{2\alpha^2u^2vw}, \quad g_2 = \frac{iv - 2(v^2-wu)}{2\alpha^2uv^2w},$$

$$g_3 = \frac{iw - 2(w^2-uv)}{2\alpha^2uvw^2}, \tag{33}$$

where $i = u+v+w$. And since these expressions hold for every equibright twofold, i.e., for every value of the brightness i, the coefficients g_1, etc. of the distance element (30) of the color threefold are herewith completely determined, independently of the relation (32) which was derived by considering the least perceptible increments of brightness. If, however, this relation is to hold, the coefficient σ, or at least the ratio σ/α, cannot be constant. In fact, substituting (33) in (32), we have

$$\alpha^2/\sigma^2 = 3 + (1/2uvw)$$
$$\times\{u^2(i-2u) + v^2(i-2v) + w^2(i-2w)\}. \tag{34}$$

If, as before, α is assumed to be *constant*, then σ, the ratio of the least perceptible brightness increment to the brightness, should be a function not only of the brightness i itself, but also of its three components u, v, w, separately. It would be interesting to test this experimentally.

The distance element (30) in the unrestricted

threefold of colors now assumes the form

$$ds^2=(1/2\alpha^2uvw)\{(i-2u+(2vw/u))du^2+\text{etc.}\},$$

which can be written

$$\alpha^2ds^2=\frac{i(du^2+dv^2+dw^2)}{2uvw}-\frac{udu^2+vdv^2+wdw^2}{uvw}+\left(\frac{du}{u}\right)^2+\left(\frac{dv}{v}\right)^2+\left(\frac{dw}{w}\right)^2. \quad (35)$$

If we had only the last triplet of terms which, with $du/u=u'$, or $u'=\log u+\text{const.}$, etc., is reduced to

$$du'^2+dv'^2+dw'^2,$$

the color threefold would be Euclidean, but owing to the presence of the first two groups of terms, not reducible to Euclidean forms, this Fechnerian threefold is distinctly *non-Euclidean* and it is, therefore, impossible to faithfully represent it by a portion of ordinary space. The color geodesics can no longer be expected to coincide with the color suites, which are expressed by linear relations between the coordinates u, v, w. More than this, a color geodesic cannot even be expected to be contained within a twofold of equibright colors ($u+v+w=\text{const.}$). That is to say, the least number of just perceptible steps from a color A to another color B, even of the same brightness i, will lead through a set of colors of different brightness. It would be of fundamental importance to verify this, qualitatively at least, by experiments.

The equations of the geodesics of the color threefold can be written down at once. For we have, as before, but with unrestricted u, v, w,

$$\frac{d}{ds}\left(g_1\frac{du}{ds}\right)-\frac{1}{2}\left[\dot{u}^2\frac{\partial g_1}{\partial u}+\dot{v}^2\frac{\partial g_2}{\partial u}+\dot{w}^2\frac{\partial g_3}{\partial u}\right]=0, \text{ etc.},$$

where dots stand for derivatives with respect to s and where, by (33),

$$\frac{\partial g_1}{\partial u}=\frac{iu-2vw}{4\alpha^2u^3vw}, \quad \frac{\partial g_1}{\partial v}=\frac{u-w}{2\alpha^2uv^2w},$$

$$\frac{\partial g_1}{\partial w}=\frac{u-v}{2\alpha^2uvw^2}, \quad \frac{\partial g_2}{\partial u}=\frac{v-w}{2\alpha^2u^2vw},$$

$$\frac{\partial g_3}{\partial u}=\frac{w-v}{2\alpha^2u^2vw}=-\frac{\partial g_2}{\partial u}, \text{ etc.}$$

Thus, and since

$$\frac{dg_1}{ds}=\frac{\partial g_1}{\partial u}\dot{u}+\frac{\partial g_1}{\partial v}\dot{v}+\frac{\partial g_1}{\partial w}\dot{w},$$

the equations of a color geodesic of the threefold become

$$[iu-2u^2+2vw]\ddot{u}=\frac{1}{4u}\Big\{(2vw-iu)\dot{u}^2+\frac{4(w-u)u^2}{v}\dot{u}\dot{v}+\frac{4(v-u)u^2}{w}\dot{u}\dot{w}+2u(v-w)(\dot{v}^2-\dot{w}^2)\Big\}, \text{ etc.} \quad (36)$$

by cyclic permutation of u, v, w. In virtue of the identity

$$g_1\dot{u}^2+g_2\dot{v}^2+g_3\dot{w}^2=1,$$

$$\text{i.e., } \dot{u}^2\left(i-2u+\frac{vw}{u}\right)+\dot{v}^2\left(i-2v+\frac{wu}{v}\right)+\dot{w}^2\left(i-2w+\frac{uv}{w}\right)=2\alpha^2uvw, \quad (36a)$$

one of the equations (36) must be a consequence of the remaining two.

In spite of the complicated form of the element ds, (35), and of the geodesic equations (36), one might suppose that it may be possible to replace u, v, w by such coordinates u', v', w' in which ds^2 is reduced to a Euclidean element, i.e. to a quadratic form with constant coefficients. The space representative of the color threefold would then be Euclidean. To decide whether this is so or not by trials (of various transformations of u, v, w) would be a hopeless task. Fortunately, there are comparatively simple and infallible criteria for the Euclidean or non-Euclidean nature of a manifold, criteria, that is, which are wholly independent of the choice of the coordinates. In the present case, of a threefold, the simplest way of deciding this question is to calculate the *principal curvatures* of the color threefold and to show that they do not vanish. As a matter of fact, they will turn out to be not only different from zero, but not even constant and, at each point (color), in general different from each other (anisotropy).

The three principal curvatures, i.e. the maxima and minima of the Riemannian curvature[10] of a threefold whose distance element is given by the quadratic form

$$ds^2=g_1du^2+g_2dv^2+g_3dw^2,$$

are

$$K_1, K_2, K_3=\frac{1}{\rho_1}, \frac{1}{\rho_2}, \frac{1}{\rho_3},$$

where ρ_1, ρ_2, ρ_3 are the roots of the cubic

$$\begin{vmatrix} (1/g_1)-b_{11}\rho & -b_{12}\rho & -b_{13}\rho \\ -b_{21}\rho & (1/g_2)-b_{22}\rho & -b_{23}\rho \\ -b_{31}\rho & -b_{32}\rho & (1/g_3)-b_{33}\rho \end{vmatrix}=0, \quad (37)$$

where, in terms of Riemann's four-index symbols,

$$b_{11}=\frac{1}{g}(23, 23), \quad b_{22}=\frac{1}{g}(31, 31), \quad b_{33}=\frac{1}{g}(12, 12),$$

$$b_{12}=\frac{1}{g}(23, 31), \quad b_{23}=\frac{1}{g}(31, 12), \quad b_{31}=\frac{1}{g}(12, 23),$$

$b_{\iota\kappa}=b_{\kappa\iota}$, and $g=g_1g_2g_3$ (the determinant of the quadratic form). The direction cosines (η_1, η_2, η_3) of the three "orientations" (surface normals) at which these curvature values are attained are given by the three linear homogeneous equations

$$\sum_{\kappa}(g^{\iota\kappa}-b_{\iota\kappa}\rho)\eta_\kappa=0 \quad (i=1, 2, 3),$$

where, in our case of orthogonal coordinates, $g^{11}=1/g_1$, etc., and $g^{12}=g^{23}=g^{31}=0$. The four-index symbols are differential expressions containing the g's and their first and second derivatives, defined by

$$(\iota\mu, \lambda\kappa)=g_\mu\left[\frac{\partial}{\partial x_\kappa}\begin{Bmatrix}\iota\mu\\ \lambda\end{Bmatrix}-\frac{\partial}{\partial x_\lambda}\begin{Bmatrix}\iota\kappa\\ \mu\end{Bmatrix}+\begin{Bmatrix}\iota\lambda\\ \beta\end{Bmatrix}\begin{Bmatrix}\beta\kappa\\ \mu\end{Bmatrix}-\begin{Bmatrix}\iota\kappa\\ \beta\end{Bmatrix}\begin{Bmatrix}\beta\lambda\\ \mu\end{Bmatrix}\right],$$

to be summed over $\beta=1, 2, 3$, where

$x_1, x_2, x_3=u, v, w$ and $\begin{Bmatrix}\iota\kappa\\ \lambda\end{Bmatrix}$ are the Christoffel symbols defined by

$$\begin{Bmatrix}\iota\kappa\\ \lambda\end{Bmatrix}=\frac{1}{2g_\lambda}\left(\frac{\partial g_{\kappa\lambda}}{\partial x_\iota}+\frac{\partial g_{\lambda\iota}}{\partial x_\kappa}-\frac{\partial g_{\iota\kappa}}{\partial x_\lambda}\right)=\begin{Bmatrix}\kappa\iota\\ \lambda\end{Bmatrix};$$

for orthogonal coordinates $g_{11}=g_1$, etc., while $g_{12}=g_{23}=g_{31}=0$. Thus, for instance,

$$(23, 23)=g_3\left[\frac{\partial}{\partial w}\begin{Bmatrix}22\\ 3\end{Bmatrix}-\frac{\partial}{\partial v}\begin{Bmatrix}23\\ 3\end{Bmatrix}+\begin{Bmatrix}22\\ \beta\end{Bmatrix}\begin{Bmatrix}\beta 3\\ 3\end{Bmatrix}-\begin{Bmatrix}23\\ \beta\end{Bmatrix}\begin{Bmatrix}\beta 2\\ 3\end{Bmatrix}\right]$$

or, since in the present case of orthogonal coordinates $\begin{Bmatrix}23\\ 1\end{Bmatrix}=0$,

$$\frac{1}{g_3}(23, 23)=\frac{\partial}{\partial w}\begin{Bmatrix}22\\ 3\end{Bmatrix}-\frac{\partial}{\partial v}\begin{Bmatrix}23\\ 3\end{Bmatrix}+\begin{Bmatrix}22\\ 1\end{Bmatrix}\begin{Bmatrix}13\\ 3\end{Bmatrix}+\begin{Bmatrix}23\\ 3\end{Bmatrix}\left[\begin{Bmatrix}22\\ 2\end{Bmatrix}-\begin{Bmatrix}23\\ 3\end{Bmatrix}\right]+\begin{Bmatrix}22\\ 3\end{Bmatrix}\left[\begin{Bmatrix}33\\ 3\end{Bmatrix}-\begin{Bmatrix}32\\ 2\end{Bmatrix}\right].$$

The developed form of the cubic (37) is, with $K\equiv 1/\rho$,

$$K^3-c_2K^2-c_1K-c_0=0, \quad (37a)$$

where

$$\begin{aligned} c_0&=b_{11}b_{22}b_{33}g_1g_2g_3=gb_{11}b_{22}b_{33},\\ c_1&=(b_{23}{}^2-b_{22}b_{33})g_2g_3\\ &\quad +\text{etc. (by cyclic permutation)},\\ c_2&=g_1b_{11}+g_2b_{22}+g_3b_{33}. \end{aligned} \quad (37b)$$

For b_{11} we have

$$g_1g_2\cdot b_{11}=-\frac{1}{2}\frac{\partial}{\partial w}\left(\frac{1}{g_3}\frac{\partial g_2}{\partial w}\right)-\frac{1}{2}\frac{\partial}{\partial v}\left(\frac{1}{g_3}\frac{\partial g_3}{\partial v}\right)-\frac{1}{4g_1g_3}\frac{\partial g_2}{\partial u}\frac{\partial g_3}{\partial u}+\frac{1}{4g_3}\frac{\partial g_3}{\partial v}\left[\frac{1}{g_2}\frac{\partial g_2}{\partial v}-\frac{1}{g_3}\frac{\partial g_3}{\partial v}\right]-\frac{1}{4g_3}\frac{\partial g_2}{\partial w}\left[\frac{1}{g_3}\frac{\partial g_3}{\partial w}-\frac{1}{g_2}\frac{\partial g_2}{\partial w}\right]$$

and, after substitution of the values (33) for g_1, g_2, g_3,

[10] See, e.g., L. Bianchi's *Geometria Differenziale*, Vol. I.

$$g_1g_2b_{11}=\frac{1}{2v^2C^2}[w(w-u)(wu-w^2+4uv)$$

$$+v^2(v-u)(i-3w)]+\frac{w(v-w)^2}{4uCA}-\frac{1}{4v^2wBC^2}$$

$$\times\{w^2(w-u)[Bw(w-u)+C(vi-3v^2+4wu)]$$

$$+(v-u)v^2[Cv(u-v)-B(4uv+wu+wv)]\},$$

where $A=iu-2u^2+2vw$, $B=iv-2v^2+2wu$,

$$C=iw-2w^2+2uv. \qquad (38)$$

The expressions for b_{22} and b_{33} can be written down from this by cyclic permutation of u, v, w and A, B, C.

Next, we have

$$g_1g_2b_{12}=\frac{\partial}{\partial u}\begin{Bmatrix}23\\3\end{Bmatrix}+\begin{Bmatrix}23\\3\end{Bmatrix}\begin{Bmatrix}13\\3\end{Bmatrix}$$

$$-\begin{Bmatrix}21\\1\end{Bmatrix}\begin{Bmatrix}13\\3\end{Bmatrix}-\begin{Bmatrix}12\\2\end{Bmatrix}\begin{Bmatrix}23\\3\end{Bmatrix}$$

$$=-\frac{1}{2}\frac{\partial}{\partial u}\left(\frac{1}{g_3}\frac{\partial g_3}{\partial v}\right)+\frac{1}{4g_3}\frac{\partial g_3}{\partial v}\left(\frac{1}{g_3}\frac{\partial g_3}{\partial u}-\frac{1}{g_2}\frac{\partial g_2}{\partial u}\right)$$

$$-\frac{1}{4g_3g_1}\frac{\partial g_3}{\partial u}\frac{\partial g_1}{\partial v}$$

and after substitution of $g_1=A/2\alpha^2u^2vw$, etc.,

$$g_1g_2b_{12}=\frac{1}{2vC^2}[2C-(w+2v)(w+2u)]$$

$$+\frac{w(w-u)}{4vBC^2}[B(w+2v)-C(v+2w)]$$

$$+\frac{w}{4AC}(w-v)(w-u). \qquad (39)$$

The values of b_{23}, b_{31} can be written down at once by cyclic permutation of u, v, w and A, B, C.

These rather complicated expressions for the b's have to be substituted in (37b), the coefficients of the cubic (37a).

It will be enough to find the roots of this cubic for the particular set of colors (forming an isochrome line)

$$u=v=w=\tfrac{1}{3}i.$$

We then have $b_{11}=b_{22}=b_{33}$, $A=B=C=iu=i^2/3$, and $g_1=g_2=g_3=27/i^2$, so that (38) gives

$$b_{11}=b_{22}=b_{33}=0$$

and (39),

$$b_{12}=b_{23}=b_{31}=-i/162.$$

Thus, by (37b),

$$c_0=0, \quad c_1=1/12i^2, \quad c_2=0,$$

and the cubic (37a) is reduced to

$$K^3-(1/12i^2)K=0.$$

Its roots are

$$K_1=0, \quad K_2=1/2(3)^{1/2}i, \quad K_3=-1/2(3)^{1/2}i. \qquad (40)$$

Such, then, are the three principal curvatures of the color threefold at any point of the isochrome line $u=v=w=i/3$. One is nil and the remaining two, numerically equal and of opposite signs, are inversely proportional to the brightness. At any rate, the three curvatures are all different and two are variable. In other words, the curvature of the Fechnerian color threefold is variable and *anisotropic*[11] as well, and the threefold cannot, therefore, be faithfully represented in ordinary Euclidean space, and not even in any space of constant curvature.

The so-called *mean curvature* at any point (color) of the threefold is defined by

$$\bar{K}=\tfrac{1}{3}(K_1+K_2+K_3).$$

Since the cubic (37a) can be written

$$K^3-c_2K^2-c_1K-c_0$$

$$=(K-K_1)(K-K_2)(K-K_3)=0,$$

we have $\bar{K}=\tfrac{1}{3}c_2$,

i.e., by (37b),

$$\bar{K}=\tfrac{1}{3}(g_1b_{11}+g_2b_{22}+g_3b_{33})$$

$$=\frac{1}{6\alpha^2uvw}\left(\frac{Ab_{11}}{u}+\frac{Bb_{22}}{v}+\frac{Cb_{33}}{w}\right), \qquad (41)$$

where b_{11}, etc. are as in (38) and $A=iu-2u^2$

[11] We may mention here that, according to a famous geometrical theorem due to F. Schur, if K were everywhere *isotropic* ($K_1=K_2=K_3$), it would also be *constant*.

$+2vw$, etc. Thus, even the mean curvature of the color threefold (which vanishes only at particular points or lines, such as $u=v=w$) is not constant, but a complicated function of u, v, w.

Hence, as has been repeatedly stated, the whole domain of colors metricized according to Fechner's law cannot be represented without distortion by a Euclidean three-dimensional model. Moreover, as we saw before, a Fechnerian color twofold of constant brightness is neither flat, nor even of constant curvature (it is not even a geodesic surface of the color threefold) and cannot, therefore, be faithfully mapped on a plane or any surface of constant curvature.

V. Bibliographic Notes

The oldest investigation on the color domain by the methods of differential geometry, and the only one known to me at the time of working out the substance of the present paper, is due to Helmholtz, who in his paper of 1891, on the *Shortest Lines in the Color System*,[12] preceded by and to a certain extent based upon his paper on the Fechner law,[13] considers as a "probable hypothesis" the distance element PP' (Deutlichkeit des Unterschiedes zweier Farben), in our notation ds,

$$ds^2=\left(\frac{du}{a+u}\right)^2+\left(\frac{dv}{b+v}\right)^2+\left(\frac{dw}{c+w}\right)^2, \quad (H, 1)$$

where a, b, c are constants and u, v, w the amounts of the fundamental colors ("Quanta der physiologischen Urfarben") contained in the color P.

Now, if this were the correct distance-element, the space representative of the color domain would be Euclidean. For, with $\xi=\log(u+a)$, etc. the form $(H, 1)$ reduces at once to

$$ds^2=d\xi^2+d\eta^2+d\zeta^2.$$

And, in fact, Helmholtz himself uses this transformation of coordinates. The color geodesics would be correctly represented by the straight lines $(\alpha\xi+\beta\eta+\gamma\zeta+\delta=0,\ \alpha'\xi+\beta'\eta+\gamma'\zeta+\delta'=0)$ of such a space,—while the color suites, which are expressed by linear equations in u, v, w, not ξ, η, ζ, would of course deviate from the geodesics. All colorimetric relations would, with Helmholtz' quadratic form, be very simple. Unfortunately, however, the form $(H, 1)$ and its implications cannot claim to satisfactorily represent the experimental facts. As we saw above, the application of the Fechner law, even with the simplifying restriction $\alpha_1=\alpha_2=\alpha_3$, leads to the much more complicated distance element (35) consisting, besides the last group of terms (to which Helmholtz' whole element is essentially reducible), of two more triads of terms not reducible to Euclidean forms. Of course, even our quadratic form (35) may turn out to be too narrow to represent the facts. One may, for instance, be driven by experimental findings to abandon the assumption of equality of the coefficients $\alpha_1, \alpha_2, \alpha_3$. Nay, even the simple Fechner law itself used in passing from the $\tilde{\omega}$'s to the α's may be found too narrow, when we would be driven back to the more general quadratic form containing $\tilde{\omega}_1, \tilde{\omega}_2, \tilde{\omega}_3$ as some functions of u, v, w to be determined empirically.

While the present paper was in the course of preparation, Dr. Loyd A. Jones of these laboratories has drawn my attention to some more recent colorimetric investigations conducted on the lines of differential geometry, namely, Schrödinger's three papers of 1920 and a paper by P. J. Bouma of 1935.[14] Space limitations do not permit us to report and discuss here all the details of these rather voluminous publications. Bouma's paper is, moreover, in the present writer's opinion, devoid of actual contributions to the subject and, in several points, quite misleading, as the reader may find out for himself by perusing it. It will suffice, therefore, to dwell at some length on the main points of Schrödinger's papers.

The main suggestion for and starting point of Schrödinger's theoretical investigations is the very concept of *brightness* (Helligkeit)[15] which he

[12] H. v. Helmholtz, *Kürzeste Linien im Farbensystem*, (Berlin Academy of Sciences, 1891) pp. 1071–1083, reprinted in *Wiss. Abhandlungen*, Vol. III, (Leipsig, Barth) pp. 460–475.

[13] *Wiss. Abh.*, III, pp. 407–459.

[14] Erwin Schrödinger, "Grundlinien einer Theorie der Farbenmetrik im Tagessehen," Ann. d. Physik, **63**, 397–426, 427–456, and 481–520 (1920). P. J. Bouma, "Grundlagen einer allgemeinen Theorie der Farbenmetrik," Amsterdam Acad. Sc., **38**, 35–45, 148–166, 258–281 (1935).

[15] Schrödinger appears to have used the term "Helligkeit" in two senses which are clearly distinguished by the nomenclature recommended by the Report of the Colori-

attempts to define *ab ovo*. Nay, he leaves it at first an open question whether such a concept can at all be consistently defined or not and shows that such a definition will be fundamentally possible if and only if a certain Pfaffian differential form, with three coordinate differentials and empirical coefficients, admits an integrating factor. After some general remarks on this question Schrödinger divides the whole field of color theory into "lower" and "higher colorimetry," the former being non-metrical, and the latter "properly metrical." The first two papers (apart from the introductory paragraphs) are wholly devoted to "lower" colorimetry and scarcely offer any essential novelty. The concept of *brightness* (Helligkeit), appears only in his third paper (dedicated to "higher" colorimetry) together with the distance element (ds) of the color threefold. Schrödinger begins by taking for ds^2 the general quadratic form (in his notation)

$$ds^2 = a_{\iota\kappa} dx_\iota dx_\kappa \qquad (S_1)$$

(to be summed over $\iota, \kappa = 1, 2, 3$), where $a_{\iota\kappa} = a_{\kappa\iota}$ are functions of the "coordinates" x_1, x_2, x_3, and by assuming that the differences of the coordinates for any two just distinguishable colors may be considered as differentials (i.e., that the $a_{\iota\kappa}$ are appreciably the same for two such colors), and that ds is numerically equal for all such pairs of colors.

In the next paragraph (§3) the concept of equibright colors and thence of brightness itself is introduced. Stated concisely, Schrödinger's definition of this concept is as follows. Let $P(x_1, x_2, x_3)$ be a given color; then a neighbor color $P'(x_1+dx_1$, etc.), of slightly different hue and saturation, is said to be *equibright* (gleich hell) with P, if, in geometrical phraseology, the element $PP'(dx_1, dx_2, dx_3)$ is *orthogonal* to the element $PQ(\delta x_1, \delta x_2, \delta x_3)$ of the isochrome line (l) passing through P.[16] Now, two such elements are said, in Riemannian geometry, to be orthogonal if

$$a_{\iota\kappa} \delta x_\iota dx_\kappa = 0 \text{ (summed over } \iota, \kappa),$$

and since, by the definition of an isochrome line, $\delta x_1, \delta x_2, \delta x_3$ are proportional to x_1, x_2, x_3,

$$a_{\iota\kappa} x_\iota dx_\kappa = 0. \qquad (S_2)$$

This, then, is, from Schrödinger's point of view, the criterion of equal brightness of the colors $P(x_1$, etc.) and $P'(x_1+dx_1$, etc.). The colors P' so defined form an infinitesimal closed line around P which can be considered as the boundary of a surface element of colors. Now, if the concept of brightness is to be admissible at all, we must claim that two colors equibright with a third should also be equibright with each other, and this condition readily leads to the requirement that all such surface elements should join up into an integral surface, and the mathematical condition for this is that the left-hand member of (S_2), a Pfaffian differential expression, should be integrable, that is to say, that there should exist an integrating factor $\mu(x_1, x_2, x_3)$ of the Pfaffian expression, such that

$$\mu a_{\iota\kappa} x_\iota dx_\kappa = dh$$

is a total differential, i.e. $\mu a_{\iota\kappa} x_\iota = \partial h / \partial x_\kappa$. Now, this imposes upon the coefficients $a_{\iota\kappa}$ a certain condition, namely, in the most general case,

$$0 = a_{\kappa 3}\left(\frac{\partial a_{\iota 1}}{\partial x_2} - \frac{\partial a_{\iota 2}}{\partial x_1}\right) x_\iota x_\kappa$$

$$+\text{etc., by cycl. permut. of } 1, 2, 3, \qquad (S_3)$$

and, in the case of an orthogonal system of coordinates ($a_{11} = g_1$, etc., $a_{12} = a_{23} = a_{31} = 0$), when (S_2) becomes

$$g_1 x_1 dx_1 + g_2 x_2 dx_2 + g_3 x_3 dx_3 = 0, \qquad (S_2')$$

$$g_1 x_1 \left(\frac{\partial g_2}{\partial x_3} - \frac{\partial g_3}{\partial x_2}\right) + \text{etc.} = 0. \qquad (S_3')$$

metry Committee of the O. S. A. [J. O. S. A. and R. S. I. **6**, 534 (1922), definition (a) and footnote 6]. Although "Helligkeit" has been translated as "brightness," the photometric concept evidently intended by Schrödinger in his quantitative discussion, yet he uses "Helligkeit" in some descriptive passages in the introspective sense where "brilliance" would be a more correct rendering. This confusion of the subjective and objective concepts of visual intensity persists to this day in German literature, where "Helligkeit" is still used without distinction between the two senses in which the terms "brilliance" and "brightness" are used in America. Similarly, "Farbton" is used by Schrödinger and others for both the introspective, subjective attribute defined in the above report as "hue" and the objective concept of monochromatic colorimetry which is called "dominant wave-length."

[16] This is equivalent to Schrödinger's original requirement that PP' should be the *shortest distance* (s) between P and the neighbor isochrome l' drawn through P'.

This necessary condition of integrability can also be proved to be a sufficient one. If, then, the g's are determined in accordance with this equation, h or any one-valued function of h which has the additive property can be considered as brightness. Schrödinger embarks upon a long-winded discussion of the condition (S_3') or the general one (S_3) and then picks out a particularly simple form of the g's which satisfies it. But it has occurred to me that all this reasoning can be avoided and that the necessary form of g_1, etc. can be directly derived from Schrödinger's condition (S_2') of equibrightness itself. In fact, from the way x_1, x_2, x_3 are introduced in his first paper one sees that these "coordinates" are identical with or proportional to the photometric brightness of the components of a color. Thus (S_2') should be identically satisfied for any dx_1, dx_2, dx_3 satisfying the condition

$$dx_1+dx_2+dx_3=0.$$

Now, eliminating dx_3 (say), we have

$$(g_1x_1-g_3x_3)dx_1+(g_2x_2-g_3x_3)dx_2=0$$

for *any* dx_1, dx_2. Hence

$$g_1x_1=g_2x_2=g_3x_3=f(x_1, x_2, x_3), \text{ say,}$$

and since $g_1x_1=(1/\mu)(\partial h/\partial x_1)$, etc., we have also $\partial h/\partial x_1=\partial h/\partial x_2=\partial h/\partial x_3$, so that h is some function of $x_1+x_2+x_3$ only and, if it be granted that the brightness of the resultant is equal to the sum of the brightnesses of the components, h can differ from $x_1+x_2+x_3$ only by a constant factor which we may as well equate to unity. Thus

$$h=x_1+x_2+x_3$$

(Schrödinger's x_1, etc. are identical with our u, v, w and $h=i$). At the same time

$$g_1=f/x_1, \quad g_2=f/x_2, \quad g_3=f/x_3.$$

Such, then, is the admissible set of the coefficients g_s. The form of the function f is restricted by (S_3') which now becomes

$$x_1\left(x_3\frac{\partial f}{\partial x_3}-x_2\frac{\partial f}{\partial x_2}\right)+x_2\left(x_1\frac{\partial f}{\partial x_1}-x_3\frac{\partial f}{\partial x_3}\right)+x_3\left(x_2\frac{\partial f}{\partial x_2}-x_1\frac{\partial f}{\partial x_1}\right)=0,$$

and this leaves yet a wide choice open for f. An obvious way of satisying this equation is to take for f any function of $h=x_1+x_2+x_3$ alone; for we then have $(df/dh)[x_1(x_3-x_2)+x_2(x_1-x_3)+x_3(x_2-x_1)]\equiv 0$. Schrödinger[17] chooses $f=1/h$, so that $g_1=1/hx_1$, etc., and the distance element of the color threefold thus becomes

$$ds^2=\frac{1}{h}\left(\frac{dx_1^2}{x_1}+\frac{dx_2^2}{x_2}+\frac{dx_3^2}{x_3}\right), \quad h=x_1+x_2+x_3, \quad (S_4)$$

as in formula (12′) of Schrödinger's third paper.

This element is much simpler than our (35), but it certainly is inacceptable. For, as Schrödinger himself admits,[18] the "constance of ds" for just perceptible differences along the spectrum of colors (tabulated on p. 498) is "very bad, almost as bad as with Helmholtz' element." In fact, the average value of $100ds$, for 17 colors of the spectrum between 640 and 480 mμ, being 1.06, one notes values as low as 0.67 and as high as 1.86. Correspondingly, the calculated and the observed $\delta\lambda$ (just perceptible wave-length increments) deviate from each other very much indeed, their differences being as large as 2.97–1.81, 1.66–2.64 mμ, etc. The discrepancy is seen emphatically from the plots of $\delta\lambda_{obs.}$ and $\delta\lambda_{calc.}$ against λ (Fig. 2 on p. 499).

In fine, Schrödinger's distance element is far from representing correctly the observed facts (spectrum).

It would seem worth while to submit to a similar test our distance element (35),[19] i.e., the coefficients $g_1=(iu-2u^2+2vw)/(2\alpha^2u^2vw)$, etc., (33). These coefficients, as I find, do not satisfy Schrödinger's integrability condition (S_3'), which, however, need not deter us from tentatively adopting them. For, Schrödinger's definition of equibrightness based on the assumption that the equibright surfaces are orthogonal to the isochromes $(g_1x_1dx_1+\text{etc.}=0)$ is by no means necessary. In other words, his condition (cf. supra) of PP' being the shortest distance between l and l' if P and P' are to be declared

[17] In order to satisfy Fechner's intensity law, with the same coefficient for all colors.

[18] Reference 14, p. 499.

[19] Such a test is now in progress with, for a first trial, Helmholtz' and Uhthoff's (1904) observed data which were used also by Schrödinger. Later on, improved data may be contemplated.

equibright, is by no means obligatory. More than this. With due respect for this now so eminent scientist, I cannot help remarking that his (1920) method of approaching the concept of "equibrightness" seems rather perverse. In fact, an equibright surface (color twofold) is most directly defined by $u+v+w=\text{const.}$, i.e. $x_1+x_2+x_3=\text{const.}$, and this gives as a relation between the dx_κ only $dx_1+dx_2+dx_3=0$, which does not at all imply the $a_{\iota\kappa}$ or our g's and does not, therefore, impose upon these coefficients any restriction. Their form, as in our treatment, can only be determined by studying the blends of $P(u, v, w)$ with the three basic colors L, M, N, through the $\bar{\omega}$'s. The form of the three $\bar{\omega}$'s may provisionally be guessed at with the aid of the Fechner law, leading to (33), but ultimately will have to be determined by observation. The correct distance element being thus constructed, it may well turn out that the isochromes are oblique, not orthogonal, to the equibright surfaces.

A Study of Harting's Criterion for Complete Achromatism*

FRED H. PERRIN†
Institute of Applied Optics, University of Rochester, Rochester, New York
(Received October 29, 1937)

If the glasses of an infinitely thin doublet obey Hartmann's three-constant dispersion formula, complete achromatism can be attained on condition that the λ_0 values of the two glasses be identical. This condition seems to have been first pointed out by Harting, and it is termed *Harting's criterion* herein. It can rarely be even approximately fulfilled in practice because the ν-values of realizable glasses would not be sufficiently different; this is shown by a plot of many current glasses and some obsolete ones. The relation of Harting's criterion to König's partial-dispersion criterion is discussed with a plot, as also is Lee's suggestion that a plot of c *versus* n_0 segregates the barium flints. Slevogt's conclusions for the more general case when a is different for the two glasses are considered briefly. Fused silica and some crystals of possible use in optical systems are also discussed; their Hartmann constants for the visible and the ultraviolet regions are tabulated. Some of these materials have favorable characteristics in the visible region, but all that were studied fail in the ultraviolet just as do glasses in the visible.

IT is well known that the achromatism of a doublet lens is not perfect; etymology to the contrary notwithstanding, the chromatic aberration of the usual doublet is as shown at A in Fig. 1. Much of Fraunhofer's short life was devoted to developing glasses that would reduce the outstanding secondary spectrum, and the task is of such importance that it was one of the first to be attacked by Abbe and Schott in their epoch-making investigations.

The amount of secondary spectrum to be expected from a given pair of glasses is usually estimated from the similarity of the partial-dispersion ratios of the two, but a parameter that was proposed by Harting[1] almost thirty years ago is perhaps worthy of more study than it has received. This parameter is a true criterion if (a) the lenses are infinitely thin and (b) both glasses obey Hartmann's dispersion formula for the same value of the exponent a. The error arising from the failure of these conditions to be completely fulfilled is negligible because only extraordinary glasses come even close to satisfying the criterion.

The focal length of a thin lens is given by the expression

$$1/f=(1/R_1-1/R_2)(n-1)=K(n-1). \quad (1)$$

For two thin lenses of index n_1 and n_2 in contact, the rate of change of focal length with wavelength is at a minimum when

$$K_1\,dn_1/d\lambda+K_2\,dn_2/d\lambda=0. \quad (2)$$

If the achromatism is to be complete (that is, if secondary spectrum is to be absent), it is obvious that this equation must be satisfied for all wavelengths, or the ratio

$$\frac{dn_1/d\lambda}{dn_2/d\lambda}=-\frac{K_2}{K_1}=\text{constant independent of }\lambda. \quad (3)$$

In other words,

$$\frac{d}{d\lambda}\left[\frac{dn_1/d\lambda}{dn_2/d\lambda}\right]=0. \quad (4)$$

Now let us assume that Hartmann's simple formula

$$n_\lambda=n_0+c/(\lambda-\lambda_0)^{a-1} \quad (5)$$

is valid, n_0, c, and λ_0 being constants that must be determined empirically for each glass. By an obvious procedure, Eq. (4) can be shown to be satisfied if

$$(\lambda_0)_1=(\lambda_0)_2. \quad (6)$$

This means that, if $a=1$ and the lenses are infinitely thin, secondary spectrum will be absent when the λ_0-values of the two glasses are identical. Inasmuch as Harting seems to have discovered this relation, we may perhaps term it for convenience *Harting's criterion*.

In 1908 Harting[2] published the Hartmann constants of the entire then-current Schott glass

* Presented in an abridged, preliminary form at the 1935 meeting of the Optical Society of America under the title, "A Simple Criterion for the Absence of Secondary Spectrum." Abstract in J. O. S. A. **26**, 230 (1936).

† Now at the Kodak Research Laboratories.

list, and he used these values for studying this criterion. This list is, of course, now obsolete in large measure, so the writer has computed the constants of the entire current Schott and Bausch and Lomb glass lists as well as those of some Chance glasses and obsolete Schott glasses. The material is too voluminous to reproduce here in its entirety, but a few values are presented in Table I.

The superiority of the achromatism that results when Eq. (1) is satisfied is shown by Fig. 1. Curve A is for an ordinary pair for which $\lambda_0=170$ and 214 mμ for the crown and the flint respectively; curve B is for a special pair for which $\lambda_0=172$ and 175 mμ, respectively.[a] A doublet made of quartz and fluorite, for which $\lambda_0=160$ and 168 mμ, respectively, has an almost flat color curve in the visible region, according to Gifford.[4]

The question now arises as to the possibility of satisfying this criterion in routine design. In selecting an achromatic pair, a considerable difference in ν-value is a desideratum, so a plot of λ_0 *versus* ν has been made in Fig. 2. To eliminate secondary spectrum, the glasses of a pair should lie on a horizontal line, but in general this is evidently impossible; the trend of the points is steeply sloped. A few glasses do have similar values of λ_0 and sufficiently dissimilar values of ν, but most of them have proved to be unstable and hence are not currently made. This instability of otherwise promising glasses has badgered pioneering glassmakers ever since Fraunhofer[5] tried to depart from the traditional crowns and flints. It is to be hoped that the limitations under which the glassmaker has labored in the past[6] have indeed been overcome as a result of experiments that are just being published.[20]

That the relation between λ_0 and ν is approximately linear was pointed out by Harting.[1] The equation he gave was

$$\lambda_0=265.0-1.536\nu. \tag{7}$$

[a] While this paper was in preparation, Malinowski (reference 3) published a paper in which the color corrections of three ordinary doublets were compared with those of three doublets made of glasses selected for the similarity of their partial-dispersion ratios. Computation shows that his special glasses fulfill Harting's criterion more nearly than the ordinary glasses, but a more detailed analysis is not warranted because his glasses admittedly have not been realized in practice.

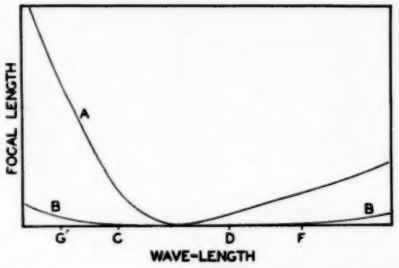


FIG. 1. Chromatic aberration curves of thin doublets. A, light silicate crown, $\lambda_0=170$ mμ, and ordinary silicate flint, $\lambda_0=214$ mμ. B, dense barium phosphate crown, $\lambda_0=172$ mμ, and borate flint, $\lambda_0=175$ mμ. The latter are the S-30 and the S-8, respectively, of Table I.

This line is drawn solid in Fig. 2. The spectral lines on which the present values are based are somewhat different from those available to Harting, so the broken line is a rather better fit. The equation of this line is

$$\lambda_0=273-1.67\nu. \tag{8}$$

The later history of Harting's criterion is interesting. Seven years after Harting, Moffitt[7] published a short paper deriving Eq. (6) without, however, mentioning Harting's work. He also considered briefly the case when λ_0 is computed from the more general equation

$$n_\lambda=n_0+c/(\lambda-\lambda_0)^{a\neq 1}, \tag{9}$$

showing that what we are calling Harting's criterion holds even when a is not equal to unity provided it is the same for both glasses. Moffitt also stated that the constant c should be very different for the two glasses to avoid great differences in the curvature of the lens surfaces. Some ten years later still, Lee[8] discussed the matter in connection with some Parsons' glasses, attributing the present Eq. (6) to Moffitt.

A more elaborate analysis from a slightly different viewpoint has recently been published by Slevogt.[9] He showed that any linear function of λ_0 (e.g., Eqs. (7) and (8)) remains linear even when λ_0 is computed from Eq. (9) instead of Eq. (5) provided a is constant for all glasses. This, of course, generalizes Moffitt's conclusion that complete achromatism can be attained by fulfilling Harting's criterion in such a case. When a is not the same for both glasses, the achromatism cannot be complete, but Slevogt

TABLES I AND II. *The Hartmann constants of some glasses and other useful materials.* For the significance of these constants, n_0, c, and λ_0, see Eq. (5) in the text; for θ, see Eq. (10); for ν', see Eq. (15). The constants of most of the glasses are computed from the published indices at the *C* line (656.28 mμ), *d* line (587.56 mμ), and *g* line (435.83 mμ) and are checked at the *F* line (486.13 mμ). The error in the fifth decimal place at the line used for checking is given in the column headed "Δn"; a positive sign means that the computed value is too high. The deviation of λ_0 from the value given by Eq. (8) is shown in the column headed "$\Delta\lambda_0$"; a positive sign means that the actual value is the higher. The deviation of θ in the third decimal place from the value given by Eq. (11) is shown in the column headed "$\Delta\theta$"; a positive sign means that the actual value is the higher. A dagger indicates that the glass is no longer made; an asterisk indicates that the entry or certain indices from which it is computed were derived from the Hartmann constants.

TABLE I. *Constants for the visible region.*

CAT. NO.	TYPE	ν	n_D	θ	n_0	c	λ_0	Δn	$\Delta\lambda_0$	$\Delta\theta$
483703	Fluor crown	70.3	1.48272	0.547	1.46687	9.7031	146.00	+ 2	− 9	− 1
PK-1	Phosphate crown	67.0	1.50447	.553	1.48756	7.3372	155.47	+ 2	− 6	0
BK-3	Borosilicate crown	65.2	1.49823	.553	1.48083	7.6217	151.55	+ 3	−12	− 4
S-30†	Barium phosphate crown	65.2	1.5760	.565	1.55705	7.9121	171.80	0	+ 8	+ 8
20-2	Borosilicate crown	64.5	1.5170	.555	1.49952	7.4111	164.86	− 1	0	− 3
BK-2	Borosilicate crown	62.0	1.50649	.562	1.48866	7.5485	166.02	+ 1	− 3	− 1
K-1	Ordinary crown	61.9	1.50970	.563	1.49194	7.4434	170.08	0	0	0
0-2388†	Telescope crown	61.7	1.5254	.568	1.50762	7.2834	179.78	− 3	+10	+ 5
BaLK-2	Light barium crown	61.1	1.51775	.564	1.49954	7.6171	171.05	− 1	0	0
160-2	Crown flint	55.1	1.5230	.570	1.50337	7.9925	182.37	− 3	+ 2	− 5
SK-9	Dense crown	55.1	1.61396	.575	1.59097	9.3072	184.27	− 2	+ 3	0
BaLF-1	Light barium flint	50.9	1.56239	.583	1.53994	8.9650	189.78	− 2	+ 2	0
S-8†	Borate flint	50.8	1.5736	.571	1.54961	9.9335	175.21	+ 1	−13	−12
KzFS-1	Telescope flint	44.0	1.61294	.584	1.58460	11.328	189.51	− 1	−10	−11
BaF-4	Barium flint	43.9	1.60550	.595	1.57847	10.491	201.19	− 2	+ 1	0
F-1	Ordinary flint	35.6	1.62572	.610	1.59259	12.474	212.89	− 4	0	0
SF-3	Dense flint	28.2	1.73976	.629	1.69227	17.266	225.82	− 6	0	+ 5
SFS-1	Extra dense flint	20.9	1.92250	.658	1.84760	25.938	243.14	−11	+ 5	+21
FORMULA	MISCELLANEOUS MATERIALS								NOTE	
LiF	Lithium fluoride (Art.)	98.4	1.39177	.558	1.38288	3.7676	161.15	−11	2	
CaF_2	Fluorite (Nat.)	95.4	1.43385	.562	1.42401	4.1529	168.13	0	1*a*	
SiO_2	Quartz (Nat.) (Ord. ray)	70.0	1.54426	.550	1.52710	7.3551	160.46	− 2	1*a*	
SiO_2	Fused silica (Art.)	67.9	1.45848	.545	1.44282	6.9605	144.35	+ 2	1*b*	
MgO	β-magnesia (Art.)	53.5	1.73789	.571	1.70744	12.865	165.33	+11	3	
KCl	Sylvite (Nat.)	44.5*	1.49038	.600*	1.46888	8.2971	203.44	−12	4	
NaCl	Rocksalt (Nat.)	42.8	1.54433	.596*	1.51936	9.7569	198.59	−11	4	
RbCl	Rubidium chloride (Art.)	42.2*	1.4922*	.582*	1.4683	9.562	188.0	−1o	5	
KBr	Potassium bromide (Art.)	31.5*	1.5581*	.566*	1.5203	15.77	172.7	+1o	6	
NaBr	Sodium bromide (Art.)	30.0*	1.6409*	.555*	1.5937	20.23	161.2	+2o	6	
KI	Potassium iodide (Art.)	21.4*	1.6634*	.558*	1.5954	28.93	164.0	+5o	6	

TABLE II. *Constants for the ultraviolet region.*

FORMULA	MATERIAL	ν'	$n_{361.05}$	n_0	c	λ_0	Δn	NOTE
LiF_2	Lithium fluoride (Art.)	21.9*	1.40133*	1.38117	4.8593	120.00	+ 1	7*a*
CaF_2	Fluorite (Nat.)	19.2	1.44534	1.42121	5.6128	128.45	−11	8*a*
SiO_2	Quartz (Nat.) (Ord. ray)	14.4	1.56347	1.52401	9.0207	132.42	−19	8*a*
SiO_2	Fused silica (Art.)	14.0	1.47511	1.44165	7.5337	135.91	−26	8*b*
MgO	β-magnesia (Art.)	9.8*	1.7733*	1.7038	14.71	149.4	−5o	9
RbCl	Rubidium chloride (Art.)	7.1*	1.5230*	1.4623	12.38	157.0	−3o	7*b*

Notes

NO.	λ_1	λ_2	λ_3	CHECK	REFERENCE
1	656.28 (Hα, *C*)	546.07 (Hg, *e*)	434.05 (Hγ, *G'*)	486.13 (Hβ, *F*)	(*a*) 4, 16. (*b*) 15, 16.
2	667.81 (He)	501.30 (H)	435.83 (Hg, *g*)	471.31 (He)	17
3	656.28 (Hα, *C*)	579.07 (Hg)	434.05 (Hγ, *G'*)	486.13 (Hβ, *F*)	10
4	670.80 (Li)	589.30 (Na, *D*)	410.18 (Hδ)	441.57 (Cd)	18
5	576.96 (Hg)	546.07 (Hg, *e*)	435.83 (Hg, *g*)	491.60 (Hg)	19
6	615.20 ‡	546.07 (Hg, *e*)	435.83 (Hg, *g*)	491.60 (Hg)	19
7	404.66 (Hg, *h*)	334.15 (Hg)	253.60 (Hg)	302.15 (Hg)	(*a*) 17. (*b*) 19.
8	404.66 (Hg, *h*)	361.05 (Cd)	257.30 (Cd)	274.86 (Cd)	(*a*) 4, 16. (*b*) 15, 16.
9	404.66 (Hg, *h*)	365.02 (Hg)	253.60 (Hg)	312.56 (Hg)	10

‡ Whether this is the faint Al 615.17 mμ line or the Hg 615.20 mμ line, which appears only under special conditions, is not stated.

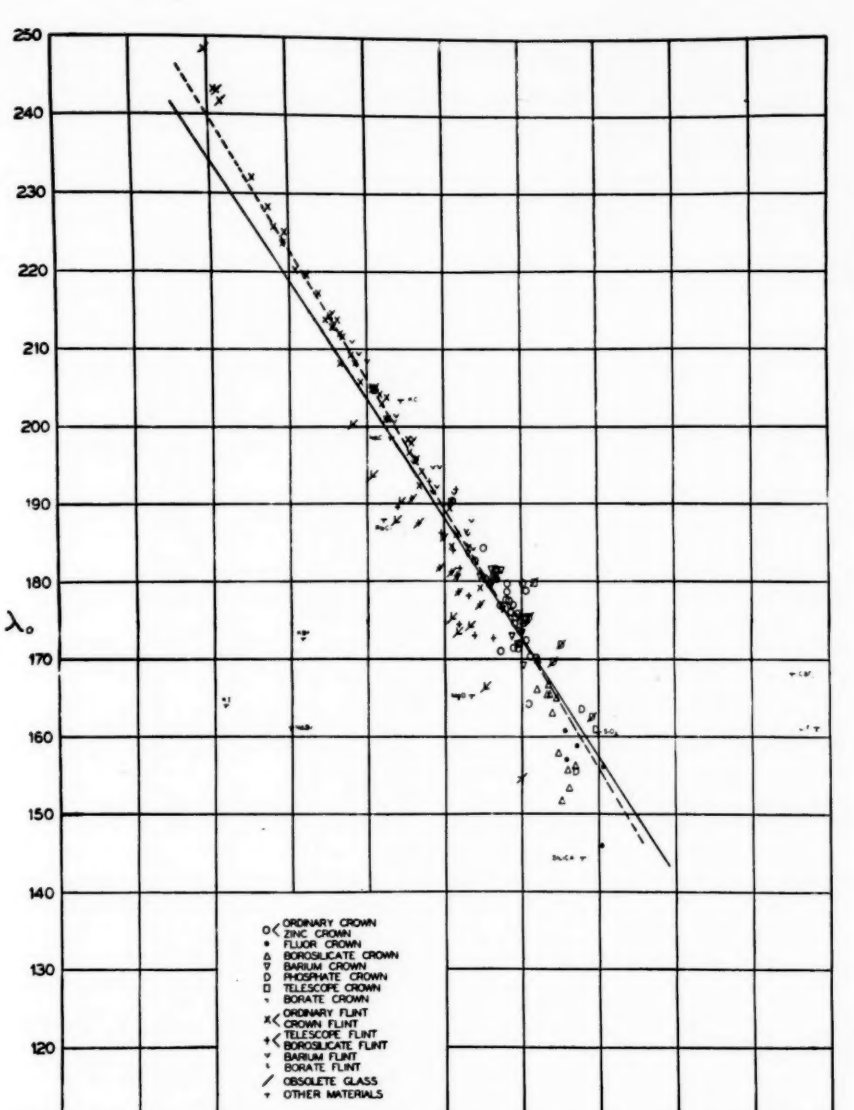


FIG. 2. Plot of λ_0 *versus* ν for the glasses included in the present study. The solid line represents Harting's formula, Eq. (7). The broken line is better representative of the present data and is expressed by Eq. (8). Values of λ_0 are in millimicrons.

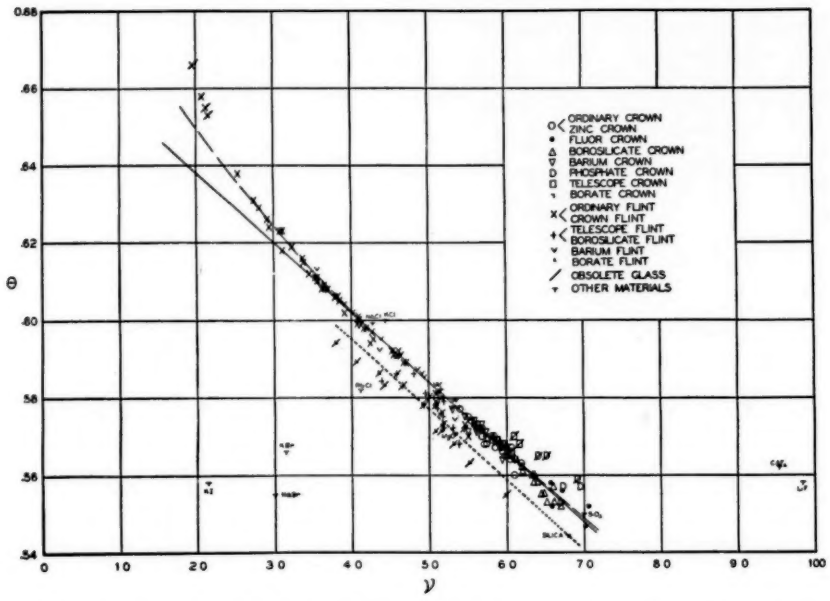


FIG. 3. Plot of θ (defined by Eq. (10)) *versus* ν. The solid line represents König's formula for ordinary glasses, Eq. (11), and the dotted line his formula for special glasses, Eq. (12). The broken line partly coinciding with the solid line is based upon the broken line in Fig. 2 and represents Eq. (14).

showed that there exists an optimum condition given by a formula that is too complicated to discuss here. In this case, the color curve of the lens exhibits the well-known maximum and minimum of focal length that is characteristic of apochromats. Therefore, instead of being satisfied with glasses for which Eq. (6) is only approximately fulfilled when λ_0 is computed from Eq. (5), Slevogt suggested selecting a pair having more favorable relations between λ_0 and a when these are computed from Eq. (9), the optimum value of a being found for each glass. Unfortunately such a procedure is tedious, and, even with the Hartmann constants available, the general criterion for optimum achromatism is not convenient for practical use. Moreover, Hartmann's formula is not sensitive to changes in the value of a. Table I shows that the errors (except for extra dense flints) are small when a is assumed to be unity throughout, and Lee's table shows that the same thing is true when a is uniformly assumed to be 1.2.[b] Since Harting's criterion, at least under present conditions, can be only approximately fulfilled, it is questionable how much would be gained by using the more elaborate method.

Referring again to Fig. 2, it will be seen that some materials other than glass are worth consideration. It is true that quartz, sylvite, and rocksalt lie near the locus of the glasses and fused silica is not well situated. On the other hand, fluorite is very favorably situated and has long been used; its great practical disadvantage is that it is no longer available in large pieces. Recently β-magnesia has been proposed for optical use, and its properties have been studied by Strong and Brice.[10] Synthetic lithium fluoride is another newcomer that might be combined with glass although it was originally proposed for use in the ultraviolet because of its high transmission. Potassium iodide, potassium bromide, and sodium bromide are also noteworthy, and a survey of the many substances listed in Landolt-Börnstein should disclose others of promise. The importance of crystalline substances has been enhanced by Stockbarger's

[b] But the value of a profoundly affects the value of λ_0, so a must be taken to be the same for all the glasses unless Slevogt's method is to be followed.

recent development of methods of growing large single crystals of lithium fluoride[11] and potassium bromide.[12]

A well-known criterion for complete achromatism described by König[13] relates to a partial-dispersion ratio, such as

$$\theta=(n_{G'}-n_F)/(n_F-n_C). \quad (10)$$

He showed that a plot of θ *versus* ν results for the most part in two straight lines, one for ordinary silicate glasses and a lower one of the same slope for special flints. Fig. 3 is such a plot of the glasses selected for the present study. König's equations for the curves were

$$\theta=0.674-0.0018\nu \quad (11)$$

and

$$\theta=0.667-0.0018\nu, \quad (12)$$

and these are represented in Fig. 3 by the solid and the dotted lines respectively. The fit of the first is clearly satisfactory. The second fits König's special glasses, but since he wrote, so many new glasses have been developed that any separation of them into two definite groups has become impossible.

In view of the similarity between Figs. 2 and 3, the question arises as to whether the relation between λ_0 and θ is linear. By substituting Eq. (5) in Eq. (10) three times, it is easy to show that

$$\theta=(\lambda_F-\lambda_{G'})(\lambda_C-\lambda_0)/(\lambda_C-\lambda_F)(\lambda_{G'}-\lambda_0). \quad (13)$$

This equation is far from linear, especially in view of the discontinuity at $\lambda_0=\lambda_{G'}$, although over the range of useful glasses, the curvature is not extreme.

Since Eq. (8) fits the points in Fig. 2 so closely, let us derive the corresponding relation between θ and ν. By combining Eqs. (8) and (13), it is easy to show that

$$\theta=(20{,}000+87.0\nu)/(27{,}400+284\nu). \quad (14)$$

This equation is plotted as the broken line in Fig. 3, and from a macroscopic standpoint, it fits the points as well as Eq. (8) fits the points in Fig. 2. To determine whether the fit is as satisfactory from a microscopic standpoint would entail an investigation of all the glasses included in this study. The deviations of some of the glasses from Eq. (11) are given in the last column of Table I. Now Eq. (11) is sensibly equivalent to Eq. (14) except for the dense flints, and the last two columns of this table correspond fairly well. Since the penultimate column gives deviations from Eq. (8), this equation apparently fits about as well as does Eq. (11) within the region wherein the latter is equivalent to Eq. (14). Pending a more detailed investigation, the lack of exact correspondence between the entries in the last two columns can be ascribed to experimental errors in the measurement of the indices. The individual discrepancies merit examination, however.

From a practical standpoint, it would appear to be immaterial whether Harting's criterion or König's be adopted. It is easiest to compute deviations from a linear relation, but Eq. (11) is an entirely satisfactory substitute for Eq. (14) in the region of ν-value in which secondary spectrum can be reduced with existing glasses. Of course, θ is far easier to compute than λ_0. Then, too, the partial-dispersion ratios of a given type of glass remain almost constant from batch to batch although the indices may vary considerably, whereas the behavior of the Hartmann constants apparently remains to be determined. On the other hand, Harting's criterion would seem to be superior in the sense of being more fundamental. To be sure, the Hartmann constants of a given material are different in the different regions of the spectrum, as a comparison of Tables I and II will show, because Eq. (5) is not accurately followed. However, λ_0 is the wavelength for which the index is theoretically infinite,

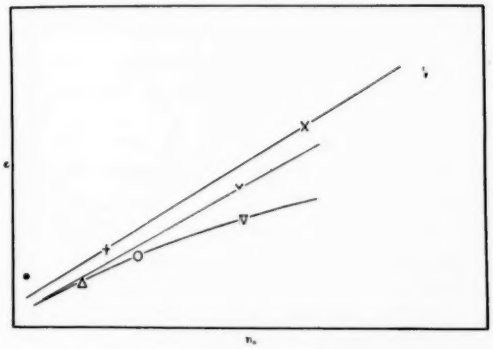


FIG. 4. Symbolic expression of Lee's hypothesis that a plot of c *versus* n_0 separates the crowns, ordinary flints, and barium flints. For the meaning of the symbols, see Fig. 5.

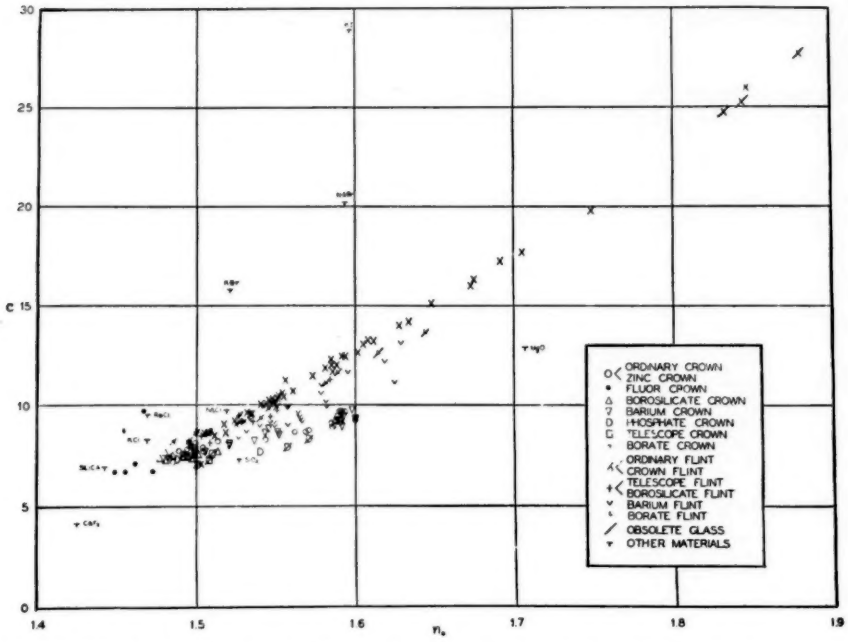


FIG. 5. Plot of c *versus* n_0 to test Lee's hypothesis expressed by Fig. 4. These glasses are the ones used for Figs. 2 and 3.

or, in other words, it indicates the location of an absorption band. Possibly some sort of correlation between the composition of a glass and its Hartmann constants might be made, thus leading to the development of new glasses having desirable properties. Lee indeed thought he had found some such correlation in regard to barium glasses when he plotted c against n_0, as shown symbolically in Fig. 4. Note that the silicate flints lie along one straight line, the crowns along a lower one, and the barium flints between. To test this hypothesis more extensively, the glasses selected for the present study were plotted in Fig. 5. There is evidently some sort of segregation in the manner suggested, although it is not sharply defined.

In view of the growing importance of the ultraviolet, the Hartmann constants of some crystals for this region are given in Table II. All the crystals of Table I are not included because some depart widely from Hartmann's law, either because of the immediate presence of an absorption band or because of inaccurate data. To determine the magnitude of the secondary spectrum of possible ultraviolet achromats, a quantity

$$\nu' = (n_{361.05} - 1)/(n_{257.30} - n_{404.66}), \qquad (15)$$

analogous to ν in the visible region, was computed for each crystal and λ_0 was plotted against it in Fig. 6. Here again the relationship is approximately linear; apparently it is as difficult to eliminate secondary spectrum in the ultraviolet as it is in the visible. In this connection, it is interesting to note that Strong and Brice went to considerable trouble to compare the residual chromatism of a β-magnesia-quartz doublet with that of a fluorite-quartz doublet.[c] Their conclusion that the latter is slightly superior in this respect could have been reached directly had Fig. 6 been available.

But the situation with respect to the ultraviolet may not be so hopeless as Fig. 6 makes it appear. Few of the materials used in optical systems obey the simple three-constant Hart-

[c] Compare a paper by Perry (reference 14) with reference 4 on the color curve of such a doublet. Perry also discusses other crystalline doublets and triplets.

mann formula well in that spectral region, at least on the basis of existing data. By applying the four-constant formula and the corresponding criterion for optimum achromatism, some combinations might appear in a more favorable light.

Despite the quarter-century that has elapsed since Harting pointed out the utility of the Hartmann formula in studying chromatic aberration, the possibilities of the method have been hardly explored. The radically different glasses just now being described should give a new impetus to the search for means of reducing residual chromatism. Harting's criterion may or may not assist in this search, but if this paper leads to an exhaustive investigation into the utility of the criterion, it will have served its purpose. One thing should be emphasized, namely, that the same principle can be applied to any dispersion formula. In the case of Hartmann's formula, a single parameter results; this is not always true. However, if a more satisfactory formula is discovered, it should be tested to determine whether an analogous criterion cannot be derived by means of it. Perchance an even more suggestive, useful, and fundamental parameter will thus be found.

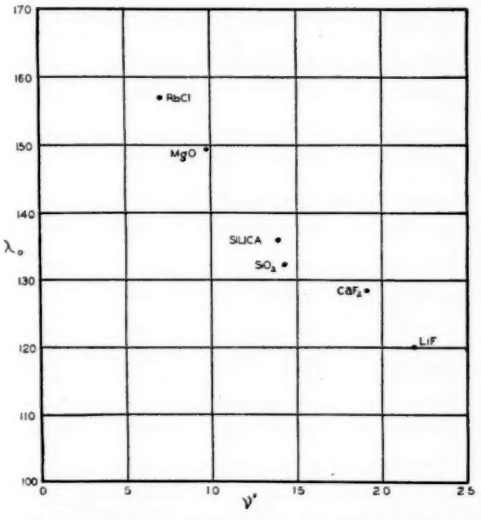


FIG. 6. Plot of λ_0 *versus* ν' for the materials listed in Table II. For the definition of ν' see Eq. (15). Compare this figure, which is for the ultraviolet, with Fig. 2, which is for the visible.

BIBLIOGRAPHY

1. H. Harting, Zeits. f. Instrumentenk. **31**, 72 (1911).
2. H. Harting, Zeits. f. Instrumentenk. **28**, 273 (1908).
3. T. Malinowski, Rev. d'optique **16**, 121 (1937).
4. J. W. Gifford, Proc. Roy. Soc. **70**, 329 (1902).
5. W. H. S. Chance, Proc. Phys. Soc. **49**, 433 (1937).
6. G. W. Morey, Sci. Mon. **42**, 541 (1936).
7. G. W. Moffitt, Phys. Rev. **11**, 144 (1918).
8. H. W. Lee, Trans. Opt. Soc. **28**, 161 (1926–7).
9. H. Slevogt, Zeits. f. Instrumentenk. **57**, 337 (1937).
10. J. Strong and R. T. Brice, J. O. S. A. **25**, 207 (1935).
11. D. C. Stockbarger, Rev. Sci. Inst. **7**, 133 (1936).
12. D. C. Stockbarger, J. O. S. A. **27**, 416 (1937).
13. See von Rohr, *The Formation of Images in Optical Instruments*, Section 199 (H. M. Stationery Office, London, 1920). Also in Czapski-Eppenstein, *Grundzüge der Theorie der Optischen Instrumente*, third edition, (Barth, 1924) pp. 295–301.
14. J. W. Perry, J. Sci. Inst. **9**, 116 (1932).
15. J. W. Gifford and W. A. Shenstone, Proc. Roy. Soc. **73**, 201 (1904).
16. J. W. Gifford, Proc. Roy. Soc. **84**, 193 (1910). Most of the indices to be found here and in references 4 and 15 are tabulated in Hardy and Perrin, *The Principles of Optics*, (McGraw-Hill, 1932) p. 394.
17. E. P. Jastram, Jr., Massachusetts Institute of Technology thesis (1935). The writer is indebted to Professor A. C. Hardy for this reference.
18. *International Critical Tables*, Vol. VII, p. 14.
19. Z. Gyulai, Zeits. f. Physik **46**, 80 (1927–8).
20. E. D. Tillyer, J. O. S. A. **28**, 1 (1938).